



**PCT**  
WELTORGANISATION FÜR GEISTIGES EIGENTUM  
Internationales Büro  
INTERNATIONALE ANMELDUNG VERÖFFENTLICHT NACH DEM VERTRAG ÜBER DIE  
INTERNATIONALE ZUSAMMENARBEIT AUF DEM GEBIET DES PATENTWESENS (PCT)

<p>(51) Internationale Patentklassifikation <sup>6</sup> : <b>C12N 15/12, C07K 14/47, 16/18, A61K 38/17, G01N 33/68, C12Q 1/68, C12N 15/62, A61K 48/00, C12N 5/10, 15/70, 15/79, 15/10</b></p>	<b>A2</b>	<p>(11) Internationale Veröffentlichungsnummer: <b>WO 99/54461</b></p> <p>(43) Internationales Veröffentlichungsdatum: 28. Oktober 1999 (28.10.99)</p>
<p>(21) Internationales Aktenzeichen: PCT/DE99/01174</p> <p>(22) Internationales Anmeldedatum: 15. April 1999 (15.04.99)</p> <p>(30) Prioritätsdaten: 198 17 948.0      17. April 1998 (17.04.98)      DE</p> <p>(71) Anmelder (für alle Bestimmungsstaaten ausser US): META-GEN GESELLSCHAFT FÜR GENOMFORSCHUNG MBH [DE/DE]; Ihnestrasse 63, D-14195 Berlin (DE).</p> <p>(72) Erfinder; und</p> <p>(75) Erfinder/Anmelder (nur für US): SPECHT, Thomas [DE/DE]; Grabenstrasse 14, D-12209 Berlin (DE). HINZMANN, Bernd [DE/DE]; Parkstrasse 19, D-13127 Berlin (DE). SCHMITT, Armin [DE/DE]; Laubacher Strasse 6/II, D-14197 Berlin (DE). PILARSKY, Christian [DE/DE]; Heinrich-Lange-Strasse 13c, D-01474 Schönfeld-Weissig (DE). DAHL, Edgar [DE/DE]; Eleonore-Procheska-Strasse 6, D-14480 Potsdam (DE). ROSENTHAL, André [DE/DE]; Koppenplatz 10, D-10115 Berlin (DE).</p>	<p>(81) Bestimmungsstaaten: JP, US, europäisches Patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE).</p> <p><b>Veröffentlicht</b> <i>Ohne internationalen Recherchenbericht und erneut zu veröffentlichen nach Erhalt des Berichts.</i></p>	
<p>(54) Title: HUMAN NUCLEIC ACID SEQUENCES OF ENDOMETRIUM TUMOUR TISSUE</p> <p>(54) Bezeichnung: MENSCHLICHE NUKLEINSÄURESEQUENZEN AUS ENDOMETRIUMTUMORGeweBE</p> <p>(57) Abstract</p> <p>The invention relates to human nucleic acid sequences (mRNA, cDNA, genomic sequences) of endometrium tumour tissue, coding for genetic products or parts thereof, in addition to the use thereof. The invention also relates to the polypeptides obtained according to said sequences and to the use thereof.</p> <p>(57) Zusammenfassung</p> <p>Es werden menschliche Nukleinsäuresequenzen-mRNA, cDNA, genomische Sequenzen – aus Endometriumtumor, die für Genprodukte oder Teile davon kodieren, und deren Verwendung beschrieben. Es werden weiterhin die über die Sequenzen erhältlichen Polypeptide und deren Verwendung beschrieben.</p>		

# **LEDIGLICH ZUR INFORMATION**

Codes zur Identifizierung von PCT-Vertragsstaaten auf den Kopfbögen der Schriften, die internationale Anmeldungen gemäss dem PCT veröffentlichen.

AL	Albanien	ES	Spanien	LS	Lesotho	SI	Slowenien
AM	Armenien	FI	Finnland	LT	Litauen	SK	Slowakei
AT	Österreich	FR	Frankreich	LU	Luxemburg	SN	Senegal
AU	Australien	GA	Gabun	LV	Lettland	SZ	Swasiland
AZ	Aserbaidshan	GB	Vereinigtes Königreich	MC	Monaco	TD	Tschad
BA	Bosnien-Herzegowina	GE	Georgien	MD	Republik Moldau	TG	Togo
BB	Barbados	GH	Ghana	MG	Madagaskar	TJ	Tadschikistan
BE	Belgien	GN	Guinea	MK	Die ehemalige jugoslawische Republik Mazedonien	TM	Turkmenistan
BF	Burkina Faso	GR	Griechenland	ML	Mali	TR	Türkei
BG	Bulgarien	HU	Ungarn	MN	Mongolei	TT	Trinidad und Tobago
BJ	Benin	IE	Irland	MR	Mauretanien	UA	Ukraine
BR	Brasilien	IL	Israel	MW	Malawi	UG	Uganda
BY	Belarus	IS	Island	MX	Mexiko	US	Vereinigte Staaten von Amerika
CA	Kanada	IT	Italien	NE	Niger	UZ	Usbekistan
CF	Zentralafrikanische Republik	JP	Japan	NL	Niederlande	VN	Vietnam
CG	Kongo	KE	Kenia	NO	Norwegen	YU	Jugoslawien
CH	Schweiz	KP	Demokratische Volksrepublik Korea	NZ	Neuseeland	ZW	Zimbabwe
CI	Côte d'Ivoire			PL	Polen		
CM	Kamerun	KR	Republik Korea	PT	Portugal		
CN	China	KZ	Kasachstan	RO	Rumänien		
CU	Kuba	LC	St. Lucia	RU	Russische Föderation		
CZ	Tschechische Republik	LI	Liechtenstein	SD	Sudan		
DE	Deutschland	LK	Sri Lanka	SE	Schweden		
DK	Dänemark	LR	Liberia	SG	Singapur		
EE	Estland						

## Menschlich Nukleinsäuresequenzen aus Endometriumtumorgewebe

5 Die Erfindung betrifft menschliche Nukleinsäuresequenzen aus Endometriumtumor, die für Genprodukte oder Teile davon kodieren, deren funktionale Gene, die mindestens ein biologisch aktives Polypeptid kodieren und deren Verwendung. Die Erfindung betrifft weiterhin die über die Sequenzen erhältlichen Polypeptide und deren Verwendung.

10 Eine Krebstodesursachen bei Frauen ist der Endometriumtumor, für dessen Bekämpfung neue Therapien notwendig sind. Bisher verwendete Therapien, wie z.B. Chemotherapie, Hormontherapie oder chirurgische Entfernung des Tumorgewebes, führen häufig nicht zu einer vollständigen Heilung.

15 Das Phänomen Krebs geht häufig einher mit der Über- oder Unterexpression gewisser Gene in den entarteten Zellen, wobei noch unklar ist, ob diese veränderten Expressionsraten Ursache oder Folge der malignen Transformation sind. Die Identifikation solcher Gene wäre ein wesentlicher Schritt für die Entwicklung neuer Therapien gegen Krebs. Der spontanen Entstehung von Krebs geht häufig eine  
20 Vielzahl von Mutationen voraus. Diese können verschiedenste Auswirkungen auf das Expressionsmuster in dem betroffenen Gewebe haben, wie z.B. Unter- oder Überexpression, aber auch Expression verkürzter Gene. Mehrere solcher Veränderungen durch solche Mutationskaskaden können schließlich zu bösartigen Entartungen führen. Die Komplexität solcher Zusammenhänge erschwert die  
25 experimentelle Herangehensweise sehr.

Für die Suche nach Kandidatengen, d.h. Genen, die im Vergleich zum Tumorgewebe im normalen Gewebe stärker exprimiert werden, wird eine Datenbank verwendet, die aus sogenannten ESTs besteht. ESTs (Expressed Sequence Tags)  
30 sind Sequenzen von cDNAs, d.h. revers transkribierten mRNAs, den Molekülen also, die die Expression von Genen widerspiegeln. Die EST-Sequenzen werden für normale und entartete Gewebe ermittelt. Solche Datenbanken werden von verschiedenen Betreibern z.T. kommerziell angeboten. Die ESTs der LifeSeq-Datenbank, die hier verwendet wird, sind in der Regel zwischen 150 und 350  
35 Nukleotide lang. Sie repräsentieren ein für ein bestimmtes Gen unverkennbares Muster, obwohl dieses Gen normalerweise sehr viel länger ist (> 2000 Nukleotide). Durch Vergleich der Expressionsmuster von normalen und Tumorgewebe können ESTs identifiziert werden, die für die Tumorentstehung und -proliferation wichtig sind. Es besteht jedoch folgendes Problem: Da durch unterschiedliche Konstruktionen der  
40 cDNA-Bibliotheken die gefundenen EST-Sequenzen zu unterschiedlichen Regionen eines unbekannten Gens gehören können, ergäbe sich in einem solchen Fall ein völlig falsches Verhältnis des Vorkommens dieser ESTs in dem jeweiligen Gewebe. Dieses würde erst bemerkt werden, wenn das vollständige Gen bekannt ist und somit die ESTs dem gleichen Gen zugeordnet werden können.

45 Es wurde nun gefunden, daß diese Fehlermöglichkeit verringert werden kann, wenn zuvor sämtliche ESTs aus dem jeweiligen Gewebstyp assembliert werden, bevor die Expressionsmuster miteinander verglichen werden. Es wurden also überlappende ESTs ein und desselben Gens zu längeren Sequenzen zusammengefaßt (s. Fig. 1, Fig. 2a und Fig.3). Durch diese Verlängerung und damit Abdeckung eines wesentlich  
50 größeren Genbereichs in jeder der jeweiligen Banken sollte der oben beschriebene Fehler weitgehendst vermieden werden. Da es hierzu keine bestehenden Softwareprodukte gab, wurden Programme für das Assemblieren von genomischen

Abschnitten verwendet, die abgewandelt eingesetzt und durch eigene Programme ergänzt wurden. Ein Flowchart der Assemblierungsprozedur ist in Fig. 2b1 – 2b4 dargestellt.

- 5 Es konnten nun die Nukleinsäure-Sequenzen Seq. ID No 1 bis Seq. ID No.141 und Seq. ID No 531-552, 554, 555 gefunden werden, die als Kandidatengene beim Endometriumtumor eine Rolle spielen.

- 10 Von besonderem Interesse sind die Nukleinsäure-Sequenzen Seq. ID Nos. 1-126 und Seq. ID No 531-552, 554, 555.

Die Erfindung betrifft somit Nukleinsäure-Sequenzen, die ein Genprodukt oder ein Teil davon kodieren, umfassend

- 15 a) eine Nukleinsäure-Sequenz, ausgewählt aus der Gruppe der Nukleinsäure-Sequenzen Seq ID Nos. 1-126 und Seq. ID No 531-552, 554, 555.
- 20 b) eine allelische Variation der unter a) genannten Nukleinsäure-Sequenzen
- oder
- 25 c) eine Nukleinsäure-Sequenz, die komplementär zu den unter a) oder b) genannten Nukleinsäure-Sequenzen ist.

- 30 Die Erfindung betrifft weiterhin eine Nukleinsäure-Sequenz gemäß einer der Sequenzen Seq ID Nos 1-126 oder eine komplementäre oder allelische Variante davon und die Nukleinsäure-Sequenzen davon, die eine 90%ige bis 95% ige Homologie zu einer humanen Nukleinsäure-Sequenz aufweisen.

- 35 Die Erfindung betrifft auch die Nukleinsäure-Sequenzen Seq. ID No. 1 bis Seq. ID No. 141 und Seq. ID No 531-552, 554, 555, die im Endometriumtumor erhöht exprimiert sind.

- 40 Die Erfindung betrifft ferner Nukleinsäure-Sequenzen, umfassend einen Teil der oben genannten Nukleinsäure-Sequenzen, in solch einer ausreichenden Größe, daß sie mit den Sequenzen Seq. ID Nos 1-126 und Seq. ID No 531-552, 554, 555 hybridisieren.

- 45 Die erfindungsgemäßen Nukleinsäure-Sequenzen weisen im allgemeinen eine Länge von mindestens 50 bis 4500 bp, vorzugsweise eine Länge von mindestens 150 bis 4000 bp, insbesondere eine Länge von 450 bis 3500 bp auf.

- 50 Mit den erfindungsgemäßen Teilsequenzen Seq. ID Nos. 1-126 und Seq. ID No 531-552, 554, 555 können gemäß gängiger Verfahrenspraxis auch Expressionskassetten konstruiert werden, wobei auf der Kassette mindestens eine der erfindungsgemäßen Nukleinsäure-Sequenzen zusammen mit mindestens einer dem Fachmann allgemein

bekannten Kontroll- oder regulatorischen Sequenz, wie z. B. einem geeigneten Promotor, kombiniert wird. Die erfindungsgemäßen Sequenzen können in sense oder antisense Orientierung eingefügt sein.

- 5 In der Literatur sind ist eine große Anzahl von Expressionskassetten bzw. Vektoren und Promotoren bekannt, die verwendet werden können.

Unter Expressionskassetten bzw. Vektoren sind zu verstehen: 1. bakterielle, wie z. B., phagescript, pBs,  $\phi$ X174, pBluescript SK, pBs KS, pNH8a, pNH16a, pNH18a, pNH46a (Stratagene), pTrc99A, pKK223-3, pKK233-3, pDR540, pRIT5 (Pharmacia),  
10 2. eukaryontische, wie z. B. pWLneo, pSV2cat, pOG44, pXT1, pSG (Stratagene), pSVK3, pBPV, pMSG, pSVL (Pharmacia).

Unter Kontroll- oder regulatorischer Sequenz sind geeignete Promotoren zu verstehen. Hierbei sind zwei bevorzugte Vektoren der pKK232-8 und der PCM7 Vektor. Im einzelnen sind folgende Promotoren gemeint: lacI, lacZ, T3, T7, gpt, lambda P<sub>R</sub>, trc, CMV, HSV Thymidin-Kinase, SV40, LTRs aus Retrovirus und Maus Metallothionein-I.

- 20 Die auf der Expressionskassette befindlichen DNA-Sequenzen können ein Fusionsprotein kodieren, das ein bekanntes Protein und ein biologisch aktives Polypeptid-Fragment umfaßt.

Die Expressionskassetten sind ebenfalls Gegenstand der vorliegenden Erfindung.

- 25 Die erfindungsgemäßen Nukleinsäure-Fragmente können zur Herstellung von Vollängen-Genen verwendet werden. Die erhältlichen Gene sind ebenfalls Gegenstand der vorliegenden Erfindung.

- 30 Die Erfindung betrifft auch die Verwendung der erfindungsgemäßen Nukleinsäure-Sequenzen, sowie die aus der Verwendung erhältlichen Gen-Fragmente.

Die erfindungsgemäßen Nukleinsäure-Sequenzen können mit geeigneten Vektoren in Wirtszellen gebracht werden, in denen als heterologer Teil die auf den  
35 Nukleinsäure-Fragmenten enthaltene genetischen Information befindet, die exprimiert wird.

Die die Nukleinsäure-Fragmente enthaltenden Wirtszellen sind ebenfalls Gegenstand der vorliegenden Erfindung.

- 40 Geeignete Wirtszellen sind z. B. prokaryontische Zellsysteme wie E. coli oder eukaryontische Zellsysteme wie tierische oder humane Zellen oder Hefen.

- 45 Die erfindungsgemäßen Nukleinsäure-Sequenzen können in sense oder antisense Form verwendet werden.

Die Herstellung der Polypeptide oder deren Fragment erfolgt durch Kultivierung der Wirtszellen gemäß gängiger Kultivierungsmethoden und anschließender Isolierung und Aufreinigung der Peptide bzw. Fragmente, ebenfalls mittels gängiger Verfahren.  
50 Die Erfindung betrifft ferner Nukleinsäure-Sequenzen, die mindestens eine Teilsequenz eines biologisch aktiven Polypeptids kodieren.

Ferner betrifft die vorliegende Erfindung Polypeptid-Teilsequenzen, sogenannte ORF (open-reading-frame)-Peptide, gemäß den Sequenzprotokollen Seq. ID Nos 142-528 und Seq. ID Nos. Seq. 561-575, 577-625, 630-635.

5

Die Erfindung betrifft ferner die Polypeptid-Sequenzen, die mindestens eine 80%ige Homologie, insbesondere eine 90%ige Homologie zu den erfindungsgemäßen Polypeptid-Teilsequenzen der ORF. ID Nos. 142-528 und Seq. ID Nos. ORF 561-575, 577-625, 630-635 aufweisen.

10

Die Erfindung betrifft auch Antikörper, die gegen ein Polypeptid oder Fragment davon gerichtete sind, welche von den erfindungsgemäßen Nukleinsäuren der Sequenzen Seq. ID No. 1 bis Seq. ID 141 und Seq. ID No 531-552, 554, 555 kodiert werden.

15

Unter Antikörper sind insbesondere monoklonale Antikörper zu verstehen.

Die erfindungsgemäßen Antikörper können u.a. durch ein Phage Display Verfahren identifiziert werden. Auch diese Antikörper sind Gegenstand der Erfindung.

20

Die erfindungsgemäßen Polypeptid-Teilsequenzen können in einem Phage Display Verfahren verwendet werden. Die mit diesem Verfahren identifizierten Polypeptide, die an die erfindungsgemäßen Polypeptid-Teilsequenzen binden, sind auch Gegenstand der Erfindung.

25

Ebenso können die erfindungsgemäßen Nukleinsäure-Sequenzen in einem Phage Display Verfahren verwendet werden.

30

Die erfindungsgemäßen Polypeptide der Sequenzen Seq. ID Nos. 142-528 und Seq. ID Nos. Seq. 561-575, 577-625, 630-635 können auch als Tool zum Auffinden von Wirkstoffen gegen den Endometriumtumor verwendet werden, was ebenfalls Gegenstand der vorliegenden Erfindung ist.

35

Ebenfalls Gegenstand der vorliegenden Erfindung ist die Verwendung der Nukleinsäure-Sequenzen gemäß den Sequenzen Seq. ID No. 1 bis Seq. ID No. 141 und Seq. ID No 531-552, 554, 555 zur Expression von Polypeptiden, die als Tools zum Auffinden von Wirkstoffen gegen den Endometriumtumor verwendet werden können.

40

Die Erfindung betrifft auch die Verwendung der gefundenen Polypeptid-Teilsequenzen Seq. ID No. 142-528 und Seq. ID Nos. Seq. 561-575, 577-625, 630-635 als Arzneimittel in der Gentherapie zur Behandlung gegen den Uterustumor, bzw. zur Herstellung eines Arzneimittels zur Behandlung gegen den Uterustumor.

45

Die Erfindung betrifft auch Arzneimittel, die mindestens eine Polypeptid-Teilsequenz Seq. ID No. 142-528 und Seq. ID Nos. Seq. 561-575, 577-625, 630-635 enthalten.

Die gefundenen erfindungsgemäßen Nukleinsäure-Sequenzen können auch genomische oder mRNA-Sequenzen sein.

50

Die Erfindung betrifft auch genomische Gene, ihre Exon- und Intronstruktur und deren Spleißvarianten, erhältlich aus den cDNAs der Sequenzen Seq. ID No. 1 bis

Seq. ID No. 141 und Seq. ID No 531-552, 554, 555, sowie deren Verwendung zusammen mit geeigneten regulativen Elementen, wie geeigneten Promotoren und/oder Enhancern.

- 5 Mit den erfindungsgemäßen Nukleinsäuren (cDNA-Sequenzen) Seq. ID No. 1-141 und Seq. ID No 531-552, 554, 555 werden genomische BAC-, PAC- und Cosmid-Bibliotheken gescreent und über komplementäre Basenpaarung (Hybridisierung) spezifisch humane Klone isoliert. Die so isolierten BAC, PAC- und Cosmid-Klone werden mit Hilfe der Fluoreszenz-in-situ-Hybridisation auf Metaphasenchromosomen
- 10 hybridisiert und entsprechende Chromosomenabschnitte identifiziert, auf denen die entsprechenden genomischen Gene liegen. BAC-, PAC- und Cosmid-Klone werden sequenziert, um die entsprechenden genomischen Gene in ihrer vollständigen Struktur (Promotoren, Enhancer, Silencer, Exons und Introns) aufzuklären. BAC-, PAC- und Cosmid-Klone können als eigenständige Moleküle für den Gentransfer
- 15 eingesetzt werden (s. Fig. 5).

Die Erfindung betrifft auch BAC-, PAC- und Cosmid-Klone, enthaltend funktionelle Gene und ihre chromosomale Lokalisation, entsprechend den Sequenzen Seq. ID. No. 1 bis Seq. ID No. 141 und Seq. ID No 531-552, 554, 555, zur Verwendung als Vehikel zum Gentransfer.

**Bedeutungen von Fachbegriffen und Abkürzungen**

5	Nukleinsäuren=	Unter Nukleinsäuren sind in der vorliegenden Erfindung zu verstehen: mRNA, partielle cDNA, vollständige cDNA und genomische Gene (Chromosomen).
	ORF =	Open Reading Frame, eine definierte Abfolge von Aminosäuren, die von der cDNA-Sequenz abgeleitet werden kann.
10	Contig =	eine Menge von DNA-Sequenzen, die aufgrund sehr großer Ähnlichkeiten zu einer Sequenz zusammengefaßt werden können (Consensus)
	Singleton=	ein Contig, der nur eine Sequenz enthält
15	Modul =	Domäne eines Proteins mit einer definierten Sequenz, die eine strukturelle Einheit darstellt und in unterschiedlichen Proteinen vorkommt
20	N =	wahlweise das Nukleotid A, T, G oder C
	X =	wahlweise eine der 20 natürlich vorkommenden Aminosäuren

**Erklärung zu den Alignmentparametern**

	minimal initial match=	minimaler anfänglicher Identitätsbereich
	maximum pads per read=	maximale Anzahl von Insertionen
30	maximum percent mismatch=	maximale Abweichung in %

**Erklärung der Abbildungen**

35	Fig. 1	zeigt die systematische Gen-Suche in der Incyte LifeSeq Datenbank.
	Fig. 2a	zeigt das Prinzip der EST-Assemblierung
40	Fig. 2b1-2b4	zeigt das gesamte Prinzip der EST-Assemblierung
	Fig. 3	zeigt die in silico Subtraktion der Genexpression in verschiedenen Geweben
45	Fig. 4a	zeigt die Bestimmung der gewebsspezifischen Expression über elektronischen Northern.
	Fig. 4b	zeigt den elektronischen Northern
50	Fig. 5	zeigt die Isolierung von genomischen BAC- und PAC-Klonen.



Die nachfolgenden Beispiele erläutern die Herstellung der erfindungsgemäßen Nukleinsäure-Sequenzen, ohne die Erfindung auf diese Beispiele und Nukleinsäure-Sequenzen zu beschränken.

5

## Beispiel 1

### Suche nach Tumor-bezogenen Kandidatengen

10

Zuerst wurden sämtliche ESTs des entsprechenden Gewebes aus der LifeSeq-Datenbank (vom Oktober 1997) extrahiert. Diese wurden dann mittels des Programms GAP4 des Staden-Pakets mit den Parametern 0% mismatch, 8 pads per read und einem minimalen match von 20 assembliert. Die nicht in die GAP4-Datenbank aufgenommenen Sequenzen (Fails) wurden erst bei 1% mismatch und dann nochmals bei 2% mismatch mit der Datenbank assembliert. Aus den Contigs der Datenbank, die aus mehr als einer Sequenz bestanden, wurden Consensussequenzen errechnet. Die Singletons der Datenbank, die nur aus einer Sequenz bestanden, wurden mit den nicht in die GAP4-Datenbank aufgenommenen Sequenzen bei 2% mismatch erneut assembliert. Wiederum wurden für die Contigs die Consensussequenzen ermittelt. Alle übrigen ESTs wurden bei 4% mismatch erneut assembliert. Die Consensussequenzen wurden abermals extrahiert und mit den vorherigen Consensussequenzen sowie den Singletons und den nicht in die Datenbank aufgenommenen Sequenzen abschließend bei 4% mismatch assembliert. Die Consensussequenzen wurden gebildet und mit den Singletons und Fails als Ausgangsbasis für die Gewebsvergleiche verwendet. Durch diese Prozedur konnte sichergestellt werden, daß unter den verwendeten Parametern sämtliche Sequenzen von einander unabhängige Genbereiche darstellten.

30 Fig. 2b1-2b4 veranschaulicht die Verlängerung der Uterusgewebe ESTs.

Die so assemblierten Sequenzen der jeweiligen Gewebe wurden anschließend mittels des gleichen Programms miteinander verglichen (Fig. 3). Hierzu wurden erst alle Sequenzen des ersten Gewebes in die Datenbank eingegeben. (Daher war es wichtig, daß diese voneinander unabhängig waren.)

Dann wurden alle Sequenzen des zweiten Gewebes mit allen des ersten verglichen. Das Ergebnis waren Sequenzen, die für das erste bzw. das zweite Gewebe spezifisch waren, sowie welche, die in beiden vorkamen. Bei Letzteren wurde das Verhältnis der Häufigkeit des Vorkommens in den jeweiligen Geweben ausgewertet. Sämtliche, die Auswertung der assemblierten Sequenzen betreffenden Programme, wurden selbst entwickelt.

Alle Sequenzen, die mehr als viermal in jeweils einem der verglichenen Gewebe vorkamen, sowie alle, die mindestens fünfmal so häufig in einem der beiden Gewebe vorkamen wurden weiter untersucht. Diese Sequenzen wurden einem elektronischen Northern (s. Beispiel 2.1) unterzogen, wodurch die Verteilung in sämtlichen Tumor- und Normal-Geweben untersucht wurde (s. Fig. 4a und Fig. 4b). Die relevanten Kandidaten wurden dann mit Hilfe sämtlicher Incyte ESTs und allen ESTs öffentlicher Datenbanken verlängert (s. Beispiel 3). Anschließend wurden die Sequenzen und ihre Übersetzung in mögliche Proteine mit allen Nukleotid- und Proteindatenbanken verglichen, sowie auf mögliche, für Proteine kodierende Regionen untersucht.

50

## B ispiel 2

### Algorithmus zur Identifikation und Verlängerung von partiellen cDNA-Sequenzen mit verändertem Expressionsmuster

Im folgenden soll ein Algorithmus zur Auffindung über- oder unterexprimierter Gene erläutert werden. Die einzelnen Schritte sind der besseren Übersicht halber auch in einem Flußdiagramm zusammengefaßt (s. Fig. 4b).

#### 2.1 Elektronischer Northern-Blot

Zu einer partiellen DNA-Sequenz S, z. B. einem einzelnen EST oder einem Contig von ESTs, werden mittels eines Standardprogramms zur Homologiesuche, z. B. BLAST (Altschul, S. F., Gish W., Miller, W., Myers, E. W. und Lipman, D. J. (1990) *J. Mol. Biol.*, **215**, 403-410), BLAST2 (Altschul, S. F., Madden, T. L., Schäffer, A. A., Zhang, J., Zhang, Z., Miller, W. und Lipman, D. J. (1997) *Nucleic Acids Research* **25** 3389-3402) oder FASTA (Pearson, W. R. und Lipman, D. J. (1988) *Proc. Natl. Acad. Sci. USA* **85** 2444-2448), die homologen Sequenzen in verschiedenen nach Geweben geordneten (privaten oder öffentlichen) EST-Bibliotheken bestimmt. Die dadurch ermittelten (relativen oder absoluten) Gewebespezifischen Vorkommenshäufigkeiten dieser Partial-Sequenz S werden als elektronischer Northern-Blot bezeichnet.

##### 2.1.1

Analog der unter 2.1 beschriebenen Verfahrensweise wurde die Sequenz Seq. ID No. 136 gefunden, die 15,6 .x stärker im Endometriumtumor als im Normalgewebe vorkommt.

Das Ergebnis ist wie folgt:

## Elektronischer Northern für SEQ. ID. NO: 57

		NORMAL	TUMOR	Verhaeltnisse	
		%Haeufigkeit	%Haeufigkeit	N/T	T/N
5	Blase	0.0078	0.0000	undef	0.0000
	Brust	0.0038	0.0075	0.5104	1.9593
	Duenn darm	0.0031	0.0000	undef	0.0000
	Eierstock	0.0000	0.0078	0.0000	undef
	Endokrines Gewebe	0.0000	0.0000	undef	undef
10	Gastrointestinal	0.0077	0.0185	0.4142	2.4145
	Gehirn	0.0096	0.0000	undef	0.0000
	Haematopoetisch	0.0000	0.0000	undef	undef
	Haut	0.0000	0.0000	undef	undef
	Hepatisch	0.0000	0.0000	undef	undef
15	Herz	0.0011	0.0000	undef	0.0000
	Hoden	0.0000	0.0117	0.0000	undef
	Lunge	0.0010	0.0000	undef	0.0000
	Magen-Speiseroehre	0.0000	0.0153	0.0000	undef
	Muskel-Skelett	0.0000	0.0000	undef	undef
20	Niere	0.0000	0.0000	undef	undef
	Pankreas	0.0000	0.0000	undef	undef
	Penis	0.0000	0.0000	undef	undef
	Prostata	0.0109	0.0085	1.2795	0.7815
	Uterus_Endometrium	0.0000	0.1583	0.0000	undef
25	Uterus_Myometrium	0.0000	0.0000	undef	undef
	Uterus_allgemein	0.0000	0.1908	0.0000	undef
	Brust-Hyperplasie	0.0032			
	Prostata-Hyperplasie	0.0089			
	Samenblase	0.0000			
30	Sinnesorgane	0.0000			
	Weisse_Blutkoerperchen	0.0000			
	Zervix	0.0106			
35		FOETUS			
		%Haeufigkeit			
	Entwicklung	0.0000			
	Gastrointestinal	0.0028			
	Gehirn	0.0000			
40	Haematopoetisch	0.0000			
	Haut	0.0000			
	Hepatisch	0.0000			
	Herz-Blutgefuesse	0.0000			
	Lunge	0.0072			
45	Nebenniere	0.0000			
	Niere	0.0124			
	Placenta	0.0000			
	Prostata	0.0000			
	Sinnesorgane	0.0000			
50					
		NORMIERTE/SUBTRAHIERTE BIBLIOTHEKEN			
		%Haeufigkeit			
	Brust	0.0068			
55	Eierstock_n	0.0000			
	Eierstock_t	0.0000			
	Endokrines_Gewebe	0.0000			
	Foetal	0.0000			
	Gastrointestinal	0.0000			
60	Haematopoetisch	0.0000			
	Haut-Muskel	0.0000			
	Hoden	0.0000			
	Lunge	0.0000			
	Nerven	0.0050			
65	Prostata	0.0000			
	Sinnesorgane	0.0000			
	Uterus_n	0.0000			

## 2.2 Fish r-Test

5 Um zu entscheiden, ob eine Partial-Sequenz *S* eines Gens in einer Bibliothek für Normal-Gewebe signifikant häufiger oder seltener vorkommt als in einer Bibliothek für entartetes Gewebe, wird Fishers Exakter Test, ein statistisches Standardverfahren (Hays, W. L., (1991) Statistics, Harcourt Brace College Publishers, Fort Worth), durchgeführt.

10 Die Null-Hypothese lautet: die beiden Bibliotheken können bezüglich der Häufigkeit zu *S* homologer Sequenzen nicht unterschieden werden. Falls die Null-Hypothese mit hinreichend hoher Sicherheit abgelehnt werden kann, wird das zu *S* gehörende Gen als interessanter Kandidat für ein Krebs-Gen akzeptiert, und es wird im nächsten Schritt versucht, eine Verlängerung seiner Sequenz zu erreichen.

15

### Beispiel 3

#### 20 Automatische Verlängerung der Partial-Sequenz

Die automatische Verlängerung der Partial-Sequenz *S* vollzieht sich in drei Schritten:

- 25 1. Ermittlung aller zu *S* homologen Sequenzen aus der Gesamtmenge der zur Verfügung stehenden Sequenzen mit Hilfe von BLAST
2. Assemblierung dieser Sequenzen mittels des Standardprogramms GAP4 (Bonfield, J. K., Smith, K. F., und Staden R. (1995), Nucleic Acids Research 23 4992-4999) (Contig-Bildung).

30

3. Berechnung einer Konsens-Sequenz *C* aus den assemblierten Sequenzen

35 Die Konsens-Sequenz *C* wird im allgemeinen länger sein als die Ausgangssequenz *S*. Ihr elektronischer Northern-Blot wird demzufolge von dem für *S* abweichen. Ein erneuter Fisher-Test entscheidet, ob die Alternativ-Hypothese der Abweichung von einer gleichmäßigen Expression in beiden Bibliotheken aufrechterhalten werden kann. Ist dies der Fall, wird versucht, *C* in gleicher Weise wie *S* zu verlängern. Diese

40 Iteration wird mit der jeweils erhaltenen Konsensus-Sequenzen  $C_i$  ( $i$ : Index der Iteration) fortgesetzt, bis die Alternativ-Hypothese verworfen wird (if  $H_0$  Exit; Abbruchkriterium I) oder bis keine automatische Verlängerung mehr möglich ist (while  $C_i > C_{i-1}$ ; Abbruchkriterium II).

45 Im Fall des Abbruchkriteriums II bekommt man mit der nach der letzten Iteration vorliegenden Konsens-Sequenz eine komplette oder annähernd komplette Sequenz eines Gens, das mit hoher statistischer Sicherheit mit Krebs in Zusammenhang gebracht werden kann.

50 Analog der oben beschriebenen Beispiele konnten die in der Tabelle I beschriebenen Nukleinsäure-Sequenzen aus Uterustumorgewebe gefunden werden.

Ferner konnten zu den einzelnen Nukleinsäure-Sequenzen die Peptidsequenzen (ORF's) bestimmt werden, die in der Tabelle II aufgelistet sind, wobei wenigen Nukleinsäure-Sequenzen kein Peptid zugeordnet werden kann und einigen Nukleinsäure-Sequenzen mehr als ein Peptid zugeordnet werden kann. Wie bereits oben erwähnt, sind sowohl die ermittelten Nukleinsäure-Sequenzen, als auch die den Nukleinsäure-Sequenzen zugeordneten Peptid-Sequenzen Gegenstand der vorliegenden Erfindung.

#### Beispiel 4

##### Kartierung der Nukleinsäure-Sequenzen auf dem humanen Genom

Die Kartierung der humanen Gene erfolgte unter Verwendung des Stanford G3 Hybrid-Panels (Stewart et al., 1997), der von Research Genetics, Huntsville, Alabama vertrieben wird. Dieses Panel besteht aus 83 verschiedenen genomischen DNAs von Mensch-Hamster Hybridzelllinien und erlaubt eine Auflösung von 500 Kilobasen. Die Hybridzelllinien wurden durch Fusion von bestrahlten diploiden menschlichen Zellen mit Zellen des Chinesischen Hamsters gewonnen. Das Rückhaltemuster der humanen Chromosomenfragmente wird mittels genspezifischer Primer in einer Polymerase-Kettenreaktion bestimmt und mit Hilfe der vom Stanford RH Server verfügbaren Software analysiert ([http://www.stanford.edu/RH/rhserver\\_form2.html](http://www.stanford.edu/RH/rhserver_form2.html)). Dieses Programm bestimmt den STS-Marker, der am nächsten zum gesuchten Gen liegt. Die entsprechende zytogenetische Bande wurde unter Verwendung des "MapView" -Programms der Genome Database (GDB), (<http://gdbwww.dkfz-heidelberg.de>) bestimmt.

Neben dem kartieren von Genen auf dem menschlichen Chromosomensatz durch verschiedene experimentelle Methoden ist es möglich die Lage von Genen auf diesem durch bioinformatische Methoden zu bestimmen. Dazu wurde das bekannte Programm e-PCR eingesetzt (Schuler GD (1998) Electronic PCR: bridging the gap between genome mapping and genome sequencing. Trends Biotechnol 16; 456-459, Schuler GD (1997). Sequence mapping by electronic PCR. Genome Res 7; 541-550). Die dabei eingesetzte Datenbank entspricht nicht mehr der in der Literatur angegebenen, sondern ist eine Weiterentwicklung, welche Daten der öffentlichen Datenbank RHdb (<http://www.ebi.ac.uk/RHdb/index.html>) einschließt. Analog zu der Kartierung durch die Hybrid-Panels erfolgte eine Auswertung der Ergebnisse mit der obengenannten Software und der Software des Whitehead-Institutes (<http://carbon.wi.mit.edu:8000/cgi-bin/contig/rhmapper.pl>).

Legende zu den Modulen:

Pfam: Protein families database of alignments and HMMs (pfam@sanger.ac.uk)

PROSITE: The PROSITE database, its status in 1999. Nucleic Acids Res. 27: 215-219 (<http://www.expasy.ch/sprot/prosite.html>)

5 TABELLE I

Sequenz ID No.:	Expression im Endometrium-Tumor:	Funktion	Module	Länge der angemeldeten Sequenz in Basen	Cytogenetische Lokalisation	nächster Marker
1	erhöht	unbekannt		1046	2p24-2p21	D2S174-D2S390
2	erhöht	Mouse mammary tumor virus proviral envelope gene Polymerase protein	2x "CSD"	373		
3	erhöht	Homolog zu Human protein kinase C-binding protein RACK17		1571	1q32.1	D1S477-D1S504
4	erhöht	Homolog zu Human mRNA for KIAA0079		1789	10q21.3-q22.2	D10S537-D10S218
5	erhöht	Caenorhabditis elegans cosmid T23B12	"BTB"	2361		
6	erhöht	Caenorhabditis elegans cosmid C01A2		1638	20q13.32-q13.33	D20S100-D20S173
7	erhöht	unbekannt		1034	12q12	D12S1589-D12S85
8	erhöht	unbekannt		947	17p11.2-p12	AFMa126yd5
9	erhöht	unbekannt		497		
10	erhöht	unbekannt		269		
11	erhöht	unbekannt		1717		
12	erhöht	unbekannt	"zf-C3HC4"	1419		
13	erhöht	unbekannt		671	2q37.3	D2S2704
14	erhöht	unbekannt		524		
15	erhöht	unbekannt		345		
16	erhöht	rGSTK1-1=glutathione S-transferase subunit 13		1060	7q33-7q36.1	WI-9353
17	erhöht	Rattus norvegicus neuritin		1721	6p23-p25.1	D6S1617-D6S1674
18	erhöht	Rattus norvegicus cytosolic NADP-dependent isocitrate dehydrogenase	"isodh"	2367	2q34	WI-1247
19	erhöht	Rat unr mRNA for unr protein with unknown function	2x "CSD"	1321	1p13.3-1q11	D1S418-D1S252
20	erhöht	Rat prostatic binding protein polypeptide c1		384		
21	erhöht	Rat GTP-binding protein (ral B)		367		
22	erhöht	R.norvegicus mRNA for TRAP-complex gamma subunit		2621	3q24-q25.2	D3S1570
23	erhöht	P.salivum mRNA for Cop1 protein	2x "G-beta"	2019	1q23.3-q24.3	D1S242-D1S416
24	erhöht	P.falciptanum pfmdr1 gene		1866	18q12.1-q12.3	AFM164ya9
25	erhöht	ORF 5' of ECLF2...ECRF3=G protein-coupled receptor homolog		1189		

Sequenz ID No.:	Expression im Endometrium-Tumor:	Funktion	Module	Länge der angemeldeten Sequenz in Basen	Cytogenetische Lokalisation	nächster Marker
26	erhöht	O.cuniculus lambda-crystallin mRNA	"3HCDH"	1418		
27	erhöht	Mus musculus flotillin		814		
28	erhöht	Mouse glycerol-3-phosphate acyltransferase		3039	10q25.1-q25.2	D10S1465
29	erhöht	Mouse clathrin-associated protein (AP47)	"Adap_comp_sub"	1448		
30	erhöht	Lycopersicon esculentum biotin-containing subunit of methylcrotonyl-CoA carboxylase	"CPSase_L_chain", "biotin_req_enzy"	1394		
31	erhöht	Leucine aminopeptidase, bovine	"Peptidase_M17"	734		
32	erhöht	Klebsiella pneumoniae possible RNA helicase (dead)	2x "DEAD"	692		
33	erhöht	Human mammaglobin Homolog	"Uteroglobin"	517		
34	erhöht	Human DNA sequence from PAC 138A5 on chromosome X		322		
35	erhöht	Human DNA sequence from clone 230G1		1559		
36	erhöht	Human DNA sequence from clone 217C2		1072		
37	erhöht	Human Cosmid Clone 26a1	"RhoGAP"	454	22.q11.21-q11.23	D22S420-D22S446
38	erhöht	Homolog zu Human chromosome 3p21.1 gene sequence		700	3p21.1	
39	erhöht	Homo sapiens DNA from chromosome 19-cosmid f21246		914		
40	erhöht	H.sapiens mRNA for Plg-1 protein		1669	17q21.31-q21.33	D17S791-D17S797
41	erhöht	H.sapiens CpG island DNA genomic Mse1 fragment		355		
42	erhöht	H.sapiens (TL5) mRNA from LNCaP cell line		2628	3q24	D3S3413
43	erhöht	Genomic sequence from Human 9q34		2535	9q34.11-q34.13	D9S179-D9S164
44	erhöht	Drosophila melanogaster misato gene	"MYB_3"	805	1q21.2	D1S305-D1S506
45	erhöht	Chicken mRNA for vitellogenin I		1279		
46	erhöht	Caenorhabditis elegans DNA from clone F31D4		1923		
47	erhöht	Caenorhabditis elegans cosmid ZK863		706		
48	erhöht	Caenorhabditis elegans cosmid ZK863		749		
49	erhöht	Caenorhabditis elegans cosmid ZK596		857	10q26.13	D10S212
50	erhöht	Caenorhabditis elegans cosmid T26A5		268		
51	erhöht	Caenorhabditis elegans cosmid T21G5		297		
52	erhöht	Caenorhabditis elegans cosmid F56D5		590		
53	erhöht	Caenorhabditis elegans cosmid F25D7		1714		
54	erhöht	Caenorhabditis elegans cosmid F08C6		1340		
55	erhöht	C.botulinum bont (partial) and nth genes		765	3q24-q23	D3S3409

Sequenz ID No.:	Expression im Endometrium-Tumor:	Funktion	Module	Länge der angemeldeten Sequenz in Basen	Cytogenetische Lokalisation	nächster Marker
56	erhöht	Bovine mRNA fragment for 49 kDa subunit of mitochondrial NADH:ubiquinone oxidoreductase (EC 1.6.5.3)	"complex1_4 9kD"	1647		
57	erhöht	Bos taurus (clone pTKD7) dopamine and cyclic AMP-regulated neuronal phosphoprotein (DARPP-32)		1166		
58	erhöht	A. thaliana mRNA for RNA helicase		487		
59	erhöht	A. thaliana glycine-rich protein (clone atGRP-4)		1630	5q23.3-q31.1	D5S396-D5S2119
60	erhöht	Saccharomyces cerevisiae Grd19p (GRD19)	2x "PX"; "BEM_DOM AIN"	1272	6q21	AFMa191wd1
61	rhöht	Saccharomyces cerevisiae chromosome XII cosmid 9328	2x "DEAD"; "helicase_C"	1914	7p12.3-p13	D7S667-D7S2427
62	erhöht	S.pombe chromosome I cosmid c13D6		608		
63		Rattus norvegicus RNA helicase with arginine-serine-rich domain		2674	17q21.31-q22	D17S797-D17S788
64		Rattus norvegicus matrilysin (MMP-7) mRNA		326		
65		Rattus norvegicus Diphor-1	2x "PDZ"	888	1q12	D1S2669-D1S498
66		Human herpesvirus-7 (HHV7) J1, G protein-coupled receptor (GCR)		202		
67		Homolog zu Human synapsin I (SYN1)		1225	1p22.3-p31.1	WI-3099
68		Homolog zu Human PAX3 gene		1093		
69		Homolog zu Human multiple exostosis 2 (EXT2)		309	1p21.3-p22.1	D1S2166
70		Homolog zu Homo sapiens integrin variant beta4E (ITGB4)		380		
71		Homolog zu Homo sapiens hCPE-R mRNA for CPE-receptor		1253		
72		Homolog zu H.sapiens mRNA for deoxyguanosine kinase		439		
73		Caenorhabditis elegans cosmid Y48E1B		1252	4p11-q12	D4S1619-D4S1600
74		Caenorhabditis elegans cosmid T21D12	"WW DO-MAIN 2"	695		
75		Caenorhabditis elegans cosmid R107		2514	13q33.3-q34	D13S261-D13S293
76		Caenorhabditis elegans cosmid M04C9		274		
77		Bovine opsin	"7tm_1"	449		
78		unbekannt		346		
79		unbekannt		1329		
80		unbekannt		805		



Sequenz ID No.:	Expression im Endometrium-Tumor:	Funktion	Module	Länge der angemeldeten Sequenz in Basen	Cytogenetische Lokalisation	nächster Marker
81		unbekannt		420		
82		unbekannt		2143	9q21.32-q22.1	D9S264-D9S257
83		unbekannt		450		
84		unbekannt		408	17q23.1-q23.2	D17S1680
85		unbekannt		311		
86		unbekannt		487		
87		unbekannt		1902	11p12-p13	WI-6150
88		unbekannt		1048	1q42.11-q43	WI-9317
89		unbekannt		804		
90		unbekannt		581		
91		unbekannt		2042		
92		unbekannt		430		
93		unbekannt		592		
94		unbekannt		674		
95		unbekannt		324		
96		unbekannt		709	5p15.33	D5S1954
97		unbekannt		562		
98		unbekannt		1948	16p13.2-p12.3	D16S499
99		unbekannt		483		
100		unbekannt		437		
101		unbekannt		359		
102		unbekannt		501		
103		unbekannt		1102	1q23.1-q23.2	D1S445-D1S431
104		unbekannt		306		
105		unbekannt		2042		
106		unbekannt		320		
107		unbekannt		506		
108		unbekannt		1276		
109		unbekannt		373		
110		unbekannt	TPR RE-PEAT™	492		
111		unbekannt		1678	6q21	D6S278-D6S302
112		unbekannt		866	9q22.1-q22.2	D9S1841-D9S196
113		unbekannt		1434	18q12.1-q12.3	D18S1124-D18S468
114		unbekannt		914	7q32.3	D7S686-D7S530
115		unbekannt		685	8p12-p11.23	D8S1821-D8S255

Sequenz ID No.:	Expression im Endometrium-Tumor:	Funktion	Module	Länge der angemeldeten Sequenz in Basen	Cyto genetische Lokalisation	nächster Marker
116		unbekannt		2646		
117		unbekannt		2667		
118		unbekannt		544		
119		unbekannt		1340	18p11.21	D18S471-D18S464
120		unbekannt		2376		
121		unbekannt		225		
122		unbekannt		1967	6q22.33-q23.1	D6S292-D6S1699
123		unbekannt		612		
124		unbekannt		1183	2q32.3-q34	D2S315-D2S2237
125		unbekannt		891	4q28.1-q31.1	
126		unbekannt		482		
127		Human triosephosphate isomerase mRNA		610		
128		Human ras inhibitor mRNA		2072	9q33.3-q34.11	
129		Human R kappa B		980		
130		Human putative interferon-related protein (SIM15)		792		
131		Human protein trafficking protein (S31iii125)	2x "EMP24-GP25L"	1092	14q32.2-14q32.33	WI-9179
132		Human protein kinase C-binding protein RACK7		1523	20q13.13-q13.2	D20S957
133		Human gene for histone H1(0)	"linker-histone"	2241	22q13.1	
134		Human cathepsin B proteinase	"Cys-protease"	631		
135		Homo sapiens cathepsin B mRNA	"Cys-protease"	980		
136		unbekannt		2238	14q24.1-14q24.3	D14S277
137		H.sapiens XG mRNA		398		
138		H.sapiens mRNA for RAB7 protein	ras	1084	7q21.3-q22.1	D7S652
139		H.sapiens mRNA for pyrroline 5-carboxylate synthetase		1259		
140		H.sapiens mRNA for beta-1,4-galactosyltransferase		1938	1q22-q23.1	
141		H.sapiens IL-13Ra		1874	Xq23	
531		Verlängerung von Seq. ID No. 19	2x "CSD"	1708	1p13.3-1q11	D1S418-D1S252
532		Verlängerung von Seq. ID No. 23	2x "G-beta"	2128	1q23.3-q24.3	D1S242-D1S416
533		Verlängerung von Seq. ID No. 25		2640		
534		Verlängerung von Seq. ID No. 32	2x "DEAD"	1245		
535		Verlängerung von Seq. ID No. 34		822		

Sequenz ID No.:	Expression im Endometrium-Tumor:	Funktion	Module	Länge der angemeldeten Sequenz in Basen	Cytogenetische Lokalisation	nächster Marker
536		Verlängerung von Seq. ID No. 43		2703	9q34.11-q34.13	D9S179-D9S164
537		Verlängerung von Seq. ID No. 44	"MYB_3"	2664	1q21.2	D1S305-D1S506
538		Verlängerung von Seq. ID No. 52		3888		
539		Verlängerung von Seq. ID No. 54		3304		
540		Verlängerung von Seq. ID No. 55		863	3q24-q23	D3S3409
541		Verlängerung von Seq. ID No. 59		1962	5q23.3-q31.1	D5S396-D5S2119
542		Verlängerung von Seq. ID No. 60	2x "PX"; "BEM_DOM AIN"	1772	6q21	AFMa191wd1
543		Verlängerung von Seq. ID No. 65	2x "PDZ"	1009	1q12	D1S2669-D1S498
544		Verlängerung von Seq. ID No. 69		2834	1p21.3-p22.1	D1S2166
545		Verlängerung von Seq. ID No. 82		2319	9q21.32-q22.1	D9S264-D9S257
546		Verlängerung von Seq. ID No. 84		2456	17q23.1-q23.2	D17S1680
547		Verlängerung von Seq. ID No. 87		2218	11p12-p13	WI-6150
548		Verlängerung von Seq. ID No. 88		2196	1q42.11-q43	WI-9317
549		Verlängerung von Seq. ID No. 93		701		
550		Verlängerung von Seq. ID No. 98		2214	16p13.2-p12.3	D16S499
551		Verlängerung von Seq. ID No. 108		1434		
552		Verlängerung von Seq. ID No. 111		2434	6q21	D6S278-D6S302
554		Verlängerung von Seq. ID No. 114		1457	7q32.3	D7S686-D7S530
555		Verlängerung von Seq. ID No. 126		741		

TABELLE II

<b>DNA-Sequenz n Seq. ID. No.</b>	<b>Peptid-S quenzen (ORF's) Seq. ID. No.</b>
<b>1</b>	<b>142</b>
	<b>143</b>
	<b>144</b>
<b>2</b>	<b>145</b>
	<b>146</b>
	<b>147</b>
<b>3</b>	<b>148</b>
	<b>149</b>
	<b>150</b>
<b>4</b>	<b>151</b>
	<b>152</b>
	<b>153</b>
<b>5</b>	<b>154</b>
	<b>155</b>
	<b>156</b>
<b>6</b>	<b>157</b>
	<b>158</b>
	<b>159</b>
<b>7</b>	<b>160</b>
	<b>161</b>
	<b>162</b>
<b>8</b>	<b>163</b>
	<b>164</b>
	<b>165</b>
<b>9</b>	<b>166</b>
	<b>167</b>
	<b>168</b>
<b>10</b>	<b>169</b>
	<b>170</b>
	<b>171</b>
<b>11</b>	<b>172</b>
	<b>173</b>
	<b>174</b>
<b>12</b>	<b>175</b>

DNA-Sequence ID. No.	Protein-Sequence (ORF's) Seq. ID. No.
12	176
	177
13	178
	179
	180
14	181
	182
	183
15	184
	185
	186
	187
16	188
	189
	190
17	191
	192
	193
18	194
	195
	196
19	197
	198
	199
20	200
	201
	202
21	203
	204
22	205
	206
	207
23	208
	209
	210
24	211

<b>DNA-Sequenzen Seq. ID. No.</b>	<b>Peptid-Sequenzen (ORF's) Seq. ID. No.</b>
<b>24</b>	<b>212</b>
	<b>213</b>
<b>25</b>	<b>214</b>
	<b>215</b>
	<b>216</b>
<b>26</b>	<b>217</b>
	<b>218</b>
	<b>219</b>
<b>27</b>	<b>220</b>
	<b>221</b>
	<b>222</b>
<b>28</b>	<b>223</b>
	<b>224</b>
	<b>225</b>
<b>29</b>	<b>226</b>
	<b>227</b>
	<b>228</b>
<b>30</b>	<b>229</b>
	<b>230</b>
	<b>231</b>
<b>31</b>	<b>232</b>
	<b>233</b>
	<b>234</b>
<b>32</b>	<b>235</b>
	<b>236</b>
	<b>237</b>
<b>33</b>	<b>238</b>
	<b>239</b>
	<b>240</b>
<b>34</b>	<b>241</b>
	<b>242</b>
	<b>243</b>
<b>35</b>	<b>244</b>
	<b>245</b>
	<b>246</b>

DNA-Sequenzen Seq. ID. No.	Peptid-Sequenzen (ORF's) Seq. ID. No.
	247
36	248
	249
37	250
	251
	252
38	253
	254
	255
39	256
	257
	258
40	259
	260
	261
41	262
	263
	264
42	265
	266
	267
43	268
	269
	270
44	271
	272
	273
45	274
	275
	276
46	277
	278
	279
47	280
	281
	282

DNA-S qu nzen S q. ID. No.	P ptid-Sequ nzen (ORF's) Seq. ID. No.
	283
48	284
	285
49	286
	287
	288
50	289
	290
	291
	292
51	293
	294
	295
52	296
	297
	298
53	299
	300
	301
54	302
	303
	304
55	305
	306
	307
56	308
	309
	310
57	311
	312
	313
58	314
	315
	316
59	317
	318



DNA-Sequenz Seq. ID. No.	Peptid-Sequenzen (ORF's) Seq. ID. No.
	319
60	320
	321
	322
61	323
	324
	325
62	326
	327
	328
63	329
	330
	331
64	332
	333
	334
	335
65	336
	337
	338
66	339
	340
	341
67	342
	343
	344
68	345
	346
	347
69	348
	349
	350
70	351
	352
	353
71	354

DNA-Sequenz n Seq. ID. N .	P ptid-Sequ nzen (ORF's) Seq. ID. No.
	355
	356
72	357
	358
	359
	360
73	361
	362
	363
74	364
	365
	366
75	367
	368
	369
76	370
	371
	372
77	373
	374
	375
78	376
	377
	378
79	379
	380
	380
	381
80	382
	383
	384
81	385
	386
	387
82	388
	389

DNA-Sequenzen S q. ID. No.	P ptid-S qu nz n (ORF's) Seq. ID. No.
	390
83	391
	392
	393
84	394
	395
85	396
	397
	398
86	399
	400
	401
	402
87	403
	404
	405
	406
88	407
	408
	409
89	410
	411
	412
90	413
	414
	415
91	416
	417
	418
92	419
	420
	421
	422
93	423
	424
	425

DNA-Sequenz n Seq. ID. No.	Peptid-S quenz n (ORF's) Seq. ID. No.
	426
94	427
	428
95	429
	430
	431
	432
96	433
	434
	435
97	436
	437
	438
98	439
	440
	441
99	442
	443
	444
100	445
	446
	447
101	448
	449
	450
102	451
	452
	453
103	454
	455
	456
104	457
	45°
	459
	460
105	461

DNA-S quenzen S q. ID. No.	P ptid-S qu n z n (ORF's) Seq. ID. No.
	462
	463
106	464
	465
107	466
	467
	468
	469
108	470
	471
	472
109	473
	474
	475
110	476
	477
	478
111	479
	480
	481
112	482
	483
	484
	485
113	486
	487
	488
114	489
	490
	491
115	492
	493
	494
	495
116	496
	497

DNA-Sequence ID. No.	Protein-Sequence (ORF's) Seq. ID. No.
	498
117	499
	500
	501
118	502
	503
	504
119	505
	506
	507
120	508
	509
	510
121	511
	512
	513
122	514
	515
	516
123	517
	518
	519
124	520
	521
	522
125	523
	524
	525
126	526
	527
	528
531	561
	562
	563
532	564
	565

DNA-Sequenz n S q. ID. No.	P ptid-S qu n z n (ORF's) Seq. ID. No.
	566
533	567
	568
	569
534	570
	571
	572
535	573
	574
	575
536	577
	578
537	579
	580
	581
538	582
	583
	584
539	585
	586
	587
540	588
	589
	590
541	591
	592
	593
542	594
	595
	596
543	597
	598
	599
544	600
	601
	602

<b>DNA-Sequenzen Seq. ID. No.</b>	<b>Peptid-Sequenz n (ORF's) Seq. ID. No.</b>
<b>545</b>	<b>603</b>
	<b>604</b>
	<b>605</b>
<b>546</b>	<b>606</b>
	<b>607</b>
	<b>608</b>
<b>547</b>	<b>609</b>
	<b>610</b>
	<b>611</b>
<b>548</b>	<b>612</b>
	<b>613</b>
	<b>614</b>
<b>549</b>	<b>615</b>
	<b>616</b>
	<b>617</b>
<b>550</b>	<b>618</b>
	<b>619</b>
	<b>620</b>
<b>551</b>	<b>621</b>
	<b>622</b>
	<b>623</b>
<b>552</b>	<b>624</b>
	<b>625</b>
<b>554</b>	<b>630</b>
	<b>631</b>
	<b>632</b>
<b>555</b>	<b>633</b>
	<b>634</b>
	<b>635</b>



Die erfinderischen Nukleinsäure-Sequenzen Seq. ID No. 1 bis Seq. ID No. 141 der ermittelten Kandidatengene und die ermittelten Aminosäure-Sequenzen Seq. ID No. 142-528 werden in dem nachfolgenden Sequenzprotokoll beschrieben.

5

### Sequenzprotokoll

#### (1) ALLGEMEINE INFORMATION:

##### 10 (i) ANMELDER:

(A) NAME: metaGen - Gesellschaft für Genomforschung mbH

(B) STRASSE: Ihnestrasse 63

(C) STADT: Berlin

(E) LAND: Deutschland

##### 15 (F) POST CODE (ZIP): D-14195

(G) TELEFON: (030)-8413 1673

(H) TELEFAX: (030)-8413 1674

##### 20 (ii) TITEL DER ERFINDUNG: Menschliche Nukleinsäure-Sequenzen aus Uterustumorgewebe

(iii) Anzahl der Sequenzen: 622

##### 25 (iv) COMPUTER READABLE FORM:

(A) MEDIUM TYPE: Floppy disk

(B) COMPUTER: IBM PC compatible

(C) OPERATING SYSTEM: PC-DOS/MS-DOS

(D) SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)

30

#### (2) INFORMATION ÜBER SEQ ID NO: :

##### (i) SEQUENZ CHARAKTERISTIK:

(A) LÄNGE: 1046 Basenpaare

##### 35 (B) TYP: Nukleinsäure

(C) STRANG: einzel

(D) TOPOLOGIE: linear

##### 40 (ii) MOLEKÜLTYP: aus einzelnen ESTs durch Assemblierung und Editierung hergestellte partielle cDNA

(iii) HYPOTHETISCH: NEIN

(iii) ANTI-SENSE: NEIN

45

##### (vi) HERKUNFT:

(A) ORGANISMUS: MENSCH

(C) ORGAN:

##### 50 (vii) SONSTIGE HERKUNFT:

(A) BIBLIOTHEK: cDNA library

(iii) ANTI-SENSE: NEIN

(vi) HERKUNFT:

5 (A) ORGANISMUS: MENSCH

(C) ORGAN:

(vii) SONSTIGE HERKUNFT:

10 (A) BIBLIOTHEK: cDNA library

(xi) SEQUENZ-BESCHREIBUNG: SEQ ID NO: 56

```
gcagccggag taagatggcg gcgctgaggg ctttgtgcgg cttccggggc gtcgcgggcc 60
aggtgctgcg gcctggggct ggagtccgat tgccgattca gccagcaga ggtgttcggc 120
15 agtggcagcc agatgtggaa tgggcacagc agtttggggg agctgttatg tacccaagca 180
aagaaacagc ccactggaag cctccacctt ggaatgatgt ggaccctcca aaggacacaa 240
ttgtgaagaa cattaccctg aactttgggc cccaacaccc agcagcgcat ggtgtcctgc 300
gactagtgat ggaattgagt ggggagatgg tgcggaagtg tgatcctcac atcgggctcc 360
tgacccgagg cactgagaag ctcatatgaat acaagacctt tcttcaggcc cttccatact 420
20 ttgaccggct agactatgtg tccatgatgt gtaacgaaca ggcctattct ctagctgtgg 480
agaagttgct aaacatccgg cctcctcctc gggcacagtg gatccgagtg ctgtttggag 540
aaatcacacg tttgttgaac cacatcatgg ctgtgaccac acatgccctg gaccttgggg 600
ccatgacccc tttcttcttg ctgtttgaag aaagggagaa gatgtttgag ttctacgagc 660
gagtgtcttg agcccgaaat catgctgctt atatccggcc aggaggagtg caccaggacc 720
25 tacccttggt gcttatggat gacatttatc agttttctaa gaacttctct cttcggcttg 780
atgagttgga ggagttgctg accaacaata ggatctggcg aaatcggaca attgacattg 840
gggtttgtaac agcagaagaa gcacttaact atggttttag tggagtgatg cttcggggct 900
caggcatcca gtgggacctg cggaagaccc agccctatga tgtttacgac caggttgagt 960
ttgatgttcc tgttggttct cgaggggact gctatgatag gtacctgtgc cgggtggagg 1020
30 agatgcgcca gtccctgaga attatcgcac agtgtctaaa caagatgcct cctggggaga 1080
tcaaggttga tgatgcaaaa gtgtctccac ctaagcgagc agagatgaag acttccatgg 1140
agtcaactgat tcatcacttt aagttgtata ctgagggcta ccaagttcct ccaggagcca 1200
catatactgc cattgaggct cccaagggag agtttggggt gtacctggtg tctgatggca 1260
gcagccgccc ttatcgatgc aagatcaagg ctctgggttt tgcccatctg gctggtttgg 1320
35 acaagatgtc taagggacac atgttggcag atgtcggtgc catcataggt acccaagata 1380
ttgtatttgg agaagtagat cggtgagcag gggagcagcg tttgatcccc cctgcctatc 1440
agcttcttct gtggagcctg ttctcactg gaaattggcc tctgtgtgtg tgtgtgtgtg 1500
tgtgtgtgtg tgtgtgtatg ttcatgtaca cttggctgtc aggccttctg tgcattgtact 1560
aaaaaaggag aaattataat aaattagccg tcttgcgccc ctaggcctaa aaaaaaaaaa 1620
40 aaaaaaaaaa aaaaaaaaaa aaaaaaa 1647
```

(2) INFORMATION ÜBER SEQ ID NO: 57:

45 (i) SEQUENZ CHARAKTERISTIK:

(A) LÄNGE: 1166 Basenpaare

(B) TYP: Nukleinsäure

(C) STRANG: einzel

(D) TOPOLOGIE: linear

50

(ii) MOLEKÜLTYP: aus einzelnen ESTs durch Assemblierung und Editierung  
hergestellte partielle cDNA

(iii) HYPOTHETISCH: NEIN

55

(iii) ANTI-SENSE: NEIN

(vi) HERKUNFT:  
 (A) ORGANISMUS: MENSCH  
 (C) ORGAN:

(vii) SONSTIGE HERKUNFT:  
 (A) BIBLIOTHEK: cDNA library

(xi) SEQUENZ-BESCHREIBUNG: SEQ ID NO: 57

```

10  cgccgcctgc gcggggggga gccagcaca gaccgccgcc gggaccccga gtcgcgcacc 60
    ccagccccac cgcccacccc gcgcgccatg gaccccaagg accgcaagaa gatccagttc 120
    tcggtgcccg cgccccttag ccagctcgac ccccgccagg tggagatgat ccggcgcagg 180
    agaccaacgc ctgccatgct gttccggctc tcagagcact cctcaccaga ggaggaagcc 240
15  tccccccacc agagagcctc aggagagggg caccatctca agtcgaagag acccaacccc 300
    tgtgcctaca caccaccttc gctgaaagct gtgcagcgca ttgctgagtc tcacctgcag 360
    tctatcagca atttgaatga gaaccaggcc tcagaggagg aggatgagct gggggagctt 420
    cgggagctgg gttatccaag agaggaagat gaggaggaag aggaggatgc agccaggctg 480
    aagtccctgaa ggtcatcagg cagtctgctg ggcaaaagac aacctgtggc cagggctctg 540
20  aagggccctg ggagcgccca cccctctgag atgagtcgga gagagatgga ggctctgagg 600
    accaagtgga agaccagca ctaagtgagc ctggggagga acctcagcgc ccttccccct 660
    ctgagcctgg cacataggca cccagcctgc atctcccagg aggaagtgga ggggacatcg 720
    ctgttcccca gaaaccact ctatcctcac cctgtttgtg gctcttcccc tcgctgcta 780
    gggctgcggc ttctgacttc tagaagacta aggctggtct gtgtttgctt gtttggccac 840
25  ctttggtgta taccagaga acctgggcac ttgctgcctg atgccacccc ctgccagtca 900
    ttctccatt caccagcgg gaggtgggat gtgagacagc ccacattgga aaatccagaa 960
    aaccgggaac agggatttgc ccttcacaat tctactcccc agatcctctc ccctggacac1020
    aggagacca cagggcagga ccctaagatc tggggaaagg aggtcctgag aaccttgagg1080
    tacccttaga tccttttcta cccacttcc tatggaggat tccaagtcaa catttgtctg1140
30  aacggcttgt aacagggttc aggttg                                     1166
  
```

(2) INFORMATION ÜBER SEQ ID NO: 58:

(i) SEQUENZ CHARAKTERISTIK:  
 (A) LÄNGE: 487 Basenpaare  
 (B) TYP: Nukleinsäure  
 (C) STRANG: einzel  
 (D) TOPOLOGIE: linear

(ii) MOLEKÜLTYP: aus einzelnen ESTs durch Assemblierung und Editierung  
 hergestellte partielle cDNA

(iii) HYPOTHETISCH: NEIN

(iii) ANTI-SENSE: NEIN

(vi) HERKUNFT:  
 (A) ORGANISMUS: MENSCH  
 (C) ORGAN:

(vii) SONSTIGE HERKUNFT:  
 (A) BIBLIOTHEK: cDNA library

(xi) SEQUENZ-BESCHREIBUNG: SEQ ID NO: 58

## (2) INFORMATION ÜBER SEQ ID NO: 310:

- 5 (A) LÄNGE: 100 Aminosäuren  
(B) TYP: Protein  
(C) STRANG: einzel  
(D) TOPOLOGIE: linear

(ii) MOLEKÜLTYP: ORF

10 (iii) HYPOTHETISCH: ja

(vi) HERKUNFT:

15 (A) ORGANISMUS: MENSCH

(xi) SEQUENZ-BESCHREIBUNG: SEQ ID NO 310:

20 SQDTMRCWVL GPKVQGNVLH NCVLWRVHII PRWRLPVGCF FAWVHNSSPK LLCPFHIWLP 60  
LPNTSAGLNR QSDSSPRPQH LGRDAPEAAQ SPQRRHLTPA 100

## (2) INFORMATION ÜBER SEQ ID NO: 311:

- 25 (A) LÄNGE: 162 Aminosäuren  
(B) TYP: Protein  
(C) STRANG: einzel  
(D) TOPOLOGIE: linear

(ii) MOLEKÜLTYP: ORF

30 (iii) HYPOTHETISCH: ja

(vi) HERKUNFT:

35 (A) ORGANISMUS: MENSCH

(xi) SEQUENZ-BESCHREIBUNG: SEQ ID NO 311:

40 RRLRGGEPT DRRDPESRT PAPPPTPRAM DPKDRKKIQF SVPAPPSQLD PRQVEMIRRR 60  
RPTPAMLFRL SEHSSPEEEA SPHQRASGEG HHLKSKRPNP CAYTPPSLKA VQRIAESHLQ120  
SISNLNENQA SEEDELGEL RELGYPREED EEEEEDAARL KS 162

## (2) INFORMATION ÜBER SEQ ID NO: 312:

- 45 (A) LÄNGE: 154 Aminosäuren  
(B) TYP: Protein  
(C) STRANG: einzel  
(D) TOPOLOGIE: linear

50 (ii) MOLEKÜLTYP: ORF

(iii) HYPOTHETISCH: ja

(vi) HERKUNFT:  
(A) ORGANISMUS: MENSCH

5 :

(xi) SEQUENZ-BESCHREIBUNG: SEQ ID NO 312:

10 VSLGRNLSAL PPLSLAHRHP ACISQEEVEG TSLFPRNPLY PHPVLCSSPR LLGLRLLTSR 60  
RLRLVCVCLF AHLWLIPREP GHLLPDAHPC QSFLHSPSGR WDVROPTLEN PENREQGFAL120  
HNSTPQILSP GHRRPTGQDP KIWGKEVLRT LRYP 154

(2) INFORMATION ÜBER SEQ ID NO: 313:

15 (A) LÄNGE: 101 Aminosäuren  
(B) TYP: Protein  
(C) STRANG: einzel  
(D) TOPOLOGIE: linear

20 (ii) MOLEKÜLTYP: ORF

(iii) HYPOTHETISCH: ja

25 (vi) HERKUNFT:  
(A) ORGANISMUS: MENSCH

:

(xi) SEQUENZ-BESCHREIBUNG: SEQ ID NO 313:

30 AQGLGLFDLR WCPSPALWW GEASSSGEEC SESRNSMAGV GLLRRIISTW RGSSWLGGAG 60  
TENWIFLRSL GSMARGVGGG AGVRDSGSRR RSVLGSPRR R 101

(2) INFORMATION ÜBER SEQ ID NO: 314:

35 (A) LÄNGE: 162 Aminosäuren  
(B) TYP: Protein  
(C) STRANG: einzel  
(D) TOPOLOGIE: linear

40 (ii) MOLEKÜLTYP: ORF

(iii) HYPOTHETISCH: ja

45 (vi) HERKUNFT:  
(A) ORGANISMUS: MENSCH

:

50 (xi) SEQUENZ-BESCHREIBUNG: SEQ ID NO 314:

SDRWTCSPPL GARSMSRFPV VAGRAPRRQE EGRSRDLQE ERLSAVCIAD REEKGCTSQE 60  
GGTTPTFPIQ QQRKKIIQAV RDNSFLIVTG NTGSGKTTQL PKYLYEAGFS QHGMIGVTQP120  
RKVA AISVAQ RVAEEMKCTL GSKVGYQVRF DDCSSKETAI KY 162

**Patentanspruch**

1. Eine Nukleinsäure-Sequenz, die ein Genprodukt oder ein Teil davon kodiert,  
umfassend
- 5 a) eine Nukleinsäure-Sequenz, ausgewählt aus der Gruppe Seq ID No 1-126 und Seq. ID No 531-552, 554, 555.
- 10 b) eine allelische Variation der unter a) genannten Nukleinsäure-Sequenzen
- oder
- 15 c) eine Nukleinsäure-Sequenz, die komplementär zu den unter a) oder b) genannten Nukleinsäure-Sequenzen ist.
2. Eine Nukleinsäure-Sequenz gemäß einer der Sequenzen Seq ID Nos 1 - 126 und Seq. ID No 531-552, 554, 555 oder eine komplementäre oder allelische Variante davon.
- 20 3. Nukleinsäure-Sequenz Seq. ID No. 1 bis Seq. ID No. 141 und Seq. ID No 531-552, 554, 555, dadurch gekennzeichnet, daß sie im Uterustumgewebe erhöht exprimiert sind.
- 25 4. BAC, PAC und Cosmid-Klone, enthaltend funktionelle Gene und ihre chromosomale Lokalisation, entsprechend den Sequenzen Seq. ID. No. 1 bis Seq. ID No. 141 und Seq. ID No 531-552, 554, 555, zur Verwendung als Vehikel zum Gentransfer.
- 30 5. Eine Nukleinsäure-Sequenz gemäß den Ansprüchen 1 bis 4, dadurch gekennzeichnet, daß sie eine 90% ige Homologie zu einer humanen Nukleinsäure-Sequenz aufweist.
- 35 6. Eine Nukleinsäure-Sequenz gemäß den Ansprüchen 1 bis 4, dadurch gekennzeichnet, daß sie eine 95% ige Homologie zu einer humanen Nukleinsäure-Sequenz aufweist.
- 40 7. Eine Nukleinsäure-Sequenz, umfassend einen Teil der in den Ansprüchen 1 bis 6 genannten Nukleinsäure-Sequenzen, in solch einer ausreichenden Größe, daß sie mit den Sequenzen gemäß den Ansprüchen 1 bis 6 hybridisieren.
- 45

8. Ein Nukleinsäure-Sequenz gemäß den Ansprüchen 1 bis 7, dadurch gekennzeichnet, daß die Größe des Fragments eine Länge von mindestens 50 bis 4500 bp aufweist.
- 5
9. Eine Nukleinsäure-Sequenz gemäß den Ansprüchen 1 bis 7, dadurch gekennzeichnet, daß die Größe des Fragments eine Länge von mindestens 50 bis 4000 bp aufweist.
- 10
10. Eine Nukleinsäure-Sequenz gemäß einem der Ansprüche 1 bis 9, die mindestens eine Teilsequenz eines biologisch aktiven Polypeptids kodiert.
- 15
11. Eine Expressionskassette, umfassend ein Nukleinsäure-Fragment oder eine Sequenz gemäß einem der Ansprüche 1 bis 9, zusammen mit mindestens einer Kontroll- oder regulatorischen Sequenz.
- 20
12. Eine Expressionskassette, umfassend ein Nukleinsäure-Fragment oder eine Sequenz gemäß Anspruch 11, worin die Kontroll- oder regulatorische Sequenz ein geeigneter Promotor ist.
- 25
13. Eine Expressionskassette gemäß einem der Ansprüche 11 und 12, dadurch gekennzeichnet, daß die auf der Kassette befindlichen DNA-Sequenzen ein Fusionsprotein kodieren, das ein bekanntes Protein und ein biologisch aktives Polypeptid-Fragment umfaßt.
- 30
14. Verwendung der Nukleinsäure-Sequenzen gemäß den Ansprüchen 1 bis 10 zur Herstellung von Vollängen-Genen.
- 35
15. Ein DNA-Fragment, umfassend ein Gen, das aus der Verwendung gemäß Anspruch 14 erhältlich ist.
- 40
16. Wirtszelle, enthaltend als heterologen Teil ihrer exprimierbaren genetischen Information ein Nukleinsäure-Fragment gemäß einem der Ansprüche 1 bis 10.
- 45
17. Wirtszelle gemäß Anspruch 16, dadurch gekennzeichnet, daß es ein prokaryontisches oder eukaryontische Zellsystem ist.
- 50
18. Wirtszelle gemäß einem der Ansprüche 16 oder 17, dadurch gekennzeichnet, daß das prokaryontische Zellsystem *E. coli* und das eukaryontische Zellsystem ein tierisches, humanes oder Hefe-Zellsystem ist.

19. Ein Verfahren zur Herstellung eines Polypeptids oder eines Fragments, dadurch gekennzeichnet, daß die Wirtszellen gemäß den Ansprüchen 16 bis 18 kultiviert werden.
- 5 20. Ein Antikörper, der gegen ein Polypeptid oder ein Fragment gerichtet ist, welches von den Nukleinsäuren der Sequenzen Seq. ID No. 1 bis Seq. ID No. 141 und Seq. ID No 531-552, 554, 555 kodiert wird, das gemäß Anspruch 19 erhältlich ist.
- 10 21. Ein Antikörper gemäß Anspruch 20, dadurch gekennzeichnet, daß er monoklonal ist.
- 15 22. Ein Antikörper gemäß Anspruch 20 dadurch gekennzeichnet, daß er ein Phage-Display-Antikörper ist.
- 20 23. Polypeptid-Teilsequenzen, gemäß den Sequenzen Seq. ID Nos. Seq. 142-528 und Seq. ID Nos. Seq. 561-575, 577-625, 630-635.
24. Polypeptid-Teilsequenzen gemäß Anspruch 23, mit mindestens 80%iger Homologie zu diesen Sequenzen.
- 25 25. Ein aus einem Phage-Display hervorgegangenes Polypeptid, welches an die Polypeptid-Teilsequenzen gemäß Anspruch 23 binden kann.
- 30 26. Polypeptid-Teilsequenzen gemäß Anspruch 23, mit mindestens 90%iger Homologie zu diesen Sequenzen.
- 35 27. Verwendung der Polypeptid-Teilsequenzen gemäß den Sequenzen Seq. ID No. 142-528 und Seq. ID Nos. Seq. 561-575, 577-625, 630-635, als Tools zum Auffinden von Wirkstoffen gegen den Uterustumor.
- 40 28. Verwendung der Nukleinsäure-Sequenzen gemäß den Sequenzen Seq. ID No. 1 bis Seq. ID No. 141 und Seq. ID No 531-552, 554, 555 zur Expression von Polypeptiden, die als Tools zum Auffinden von Wirkstoffen gegen den Endometriumtumor verwendet werden können.
- 45 29. Verwendung der Nukleinsäure-Sequenzen Seq. ID No. 1 bis Seq. ID No. 141 und Seq. ID No 531-552, 554, 555 in sense oder antisense Form.



30. Verwendung der Polypeptid-Teilsequenzen Seq. ID No. 142-528 und Seq. ID Nos. Seq. 561-575, 577-625, 630-635 als Arzneimittel in der Gentherapie zur Behandlung des Endometriumtumor.
- 5 31. Verwendung der Polypeptid-Teilsequenzen Seq. ID No. 142-528 und Seq. ID Nos. Seq. 561-575, 577-625, 630-635, zur Herstellung eines Arzneimittels zur Behandlung gegen den Endometriumtumor.
- 10 32. Arzneimittel, enthaltend mindestens eine Polypeptid-Teilsequenz Seq. ID No. 142-528 und Seq. ID Nos. Seq. 561-575, 577-625, 630-635.
- 15 33. Eine Nukleinsäure-Sequenz gemäß den Ansprüchen 1 bis 10, dadurch gekennzeichnet, daß es eine genomische Sequenz ist.
- 20 34. Eine Nukleinsäure-Sequenz gemäß den Ansprüchen 1 bis 10, dadurch gekennzeichnet, daß es eine mRNA-Sequenz ist.
- 25 35. Genomische Gene, ihre Promotoren, Enhancer, Silencer, Exonstruktur, Intronstruktur und deren Spleißvarianten, erhältlich aus den cDNAs der Sequenzen Seq. ID No. 1 bis Seq. ID No. 141 und Seq. ID No 531-552, 554, 555.
- 30 36. Verwendung der genomischen Gene gemäß Anspruch 35, zusammen mit geeigneten regulativen Elementen.
37. Verwendung gemäß Anspruch 36, dadurch gekennzeichnet, daß das regulative Element ein geeigneter Promotor und/ oder Enhancer ist.
- 35 38. Eine Nukleinsäure-Sequenz gemäß den Ansprüchen 1 bis 7, dadurch gekennzeichnet, daß die Größe des Fragments eine Länge von mindestens 300 bis 3500 bp aufweist.

1/10

## Systematische Gen-Suche in der Incyte LifeSeq Datenbank

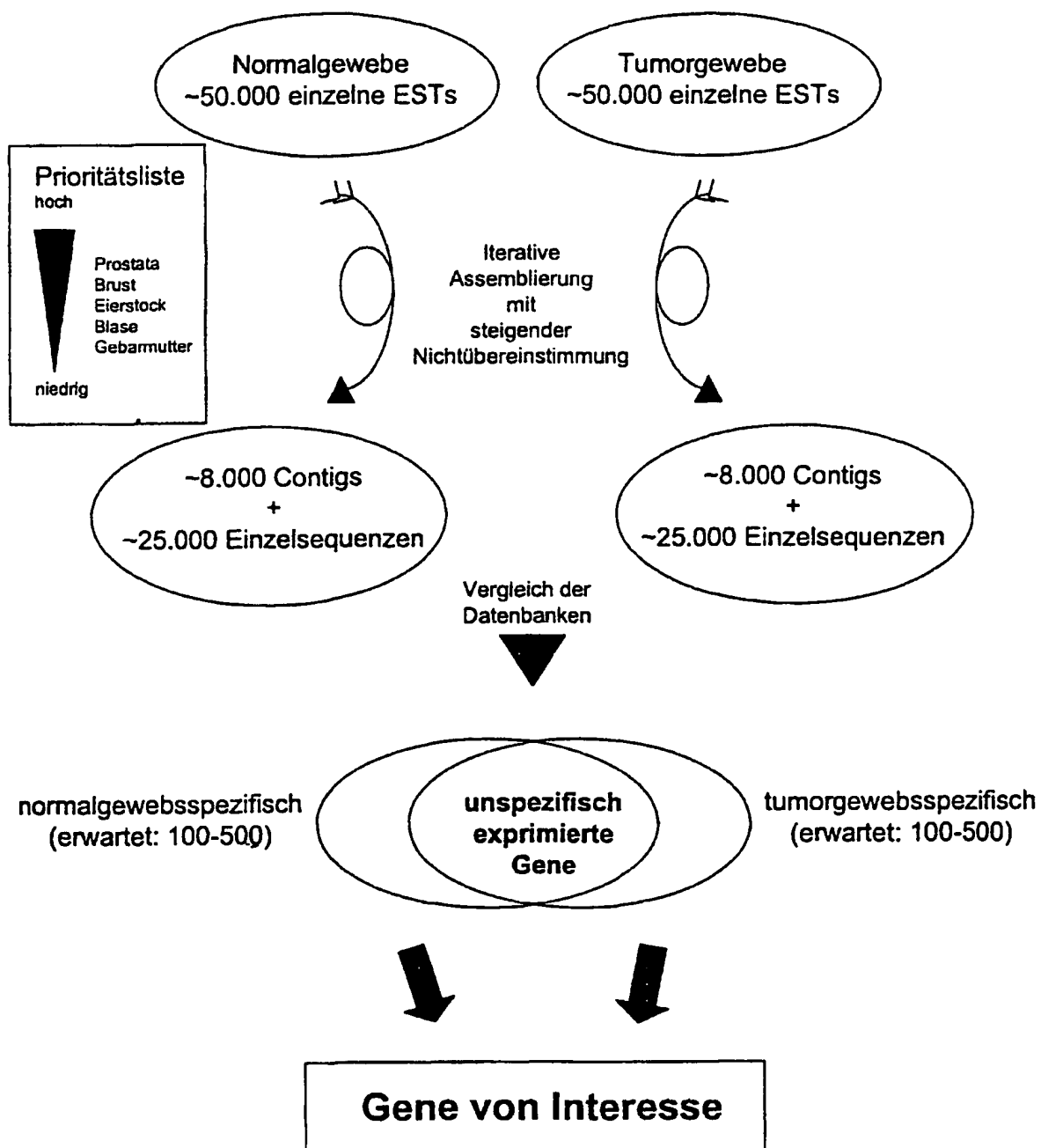


Fig. 1

2/10

# Prinzip der EST-Assemblierung

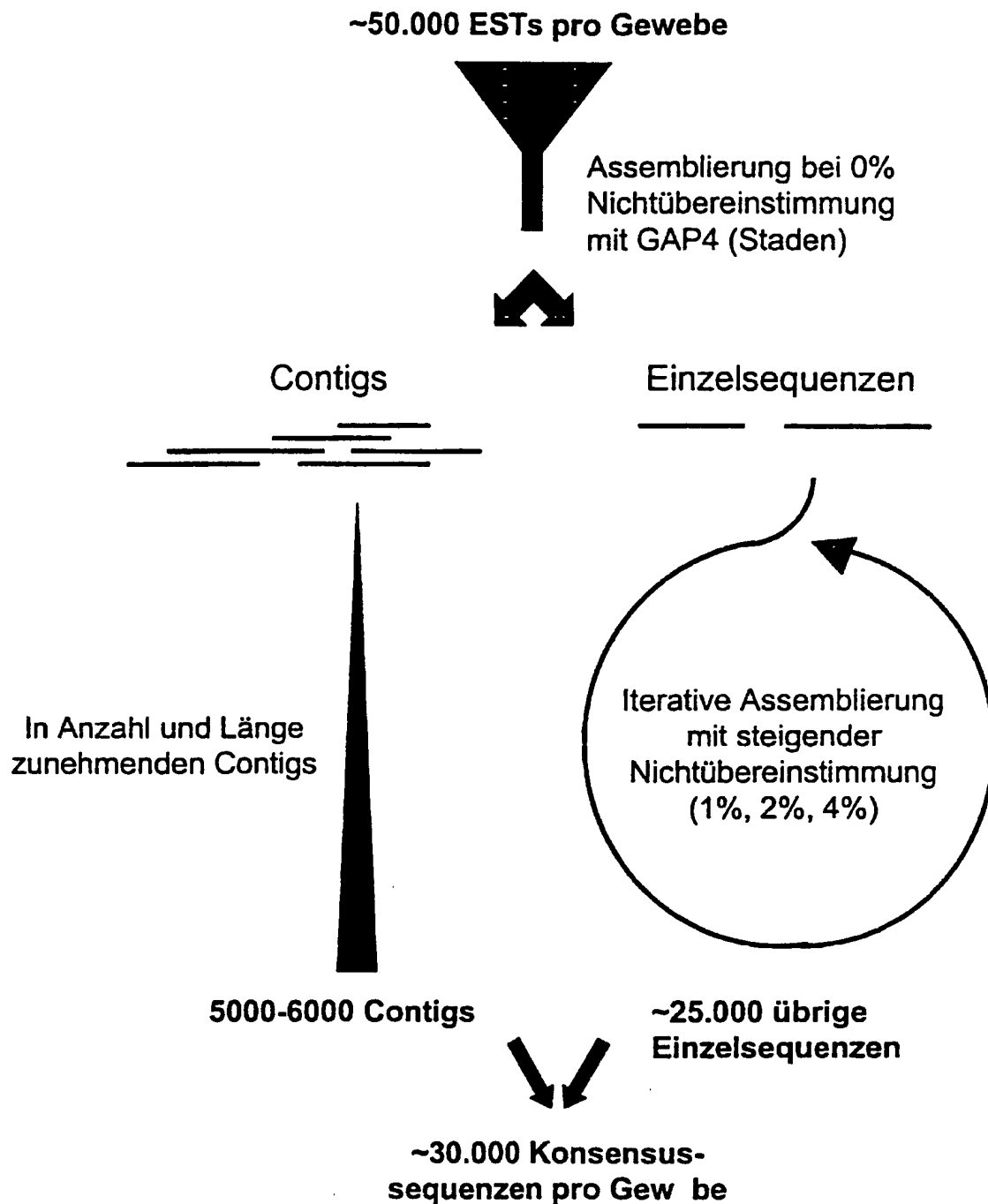


Fig. 2a

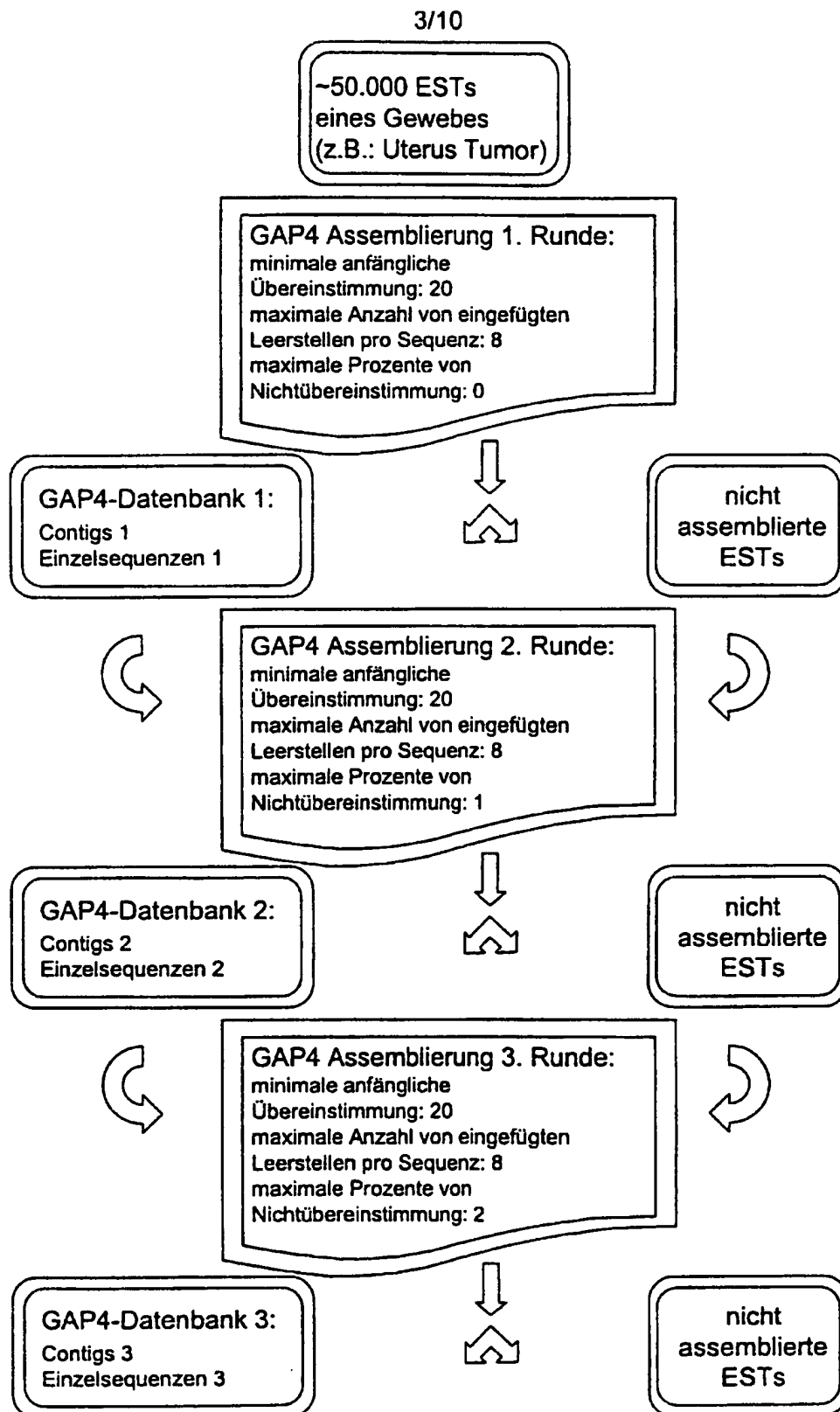


Fig. 2b1

4/10

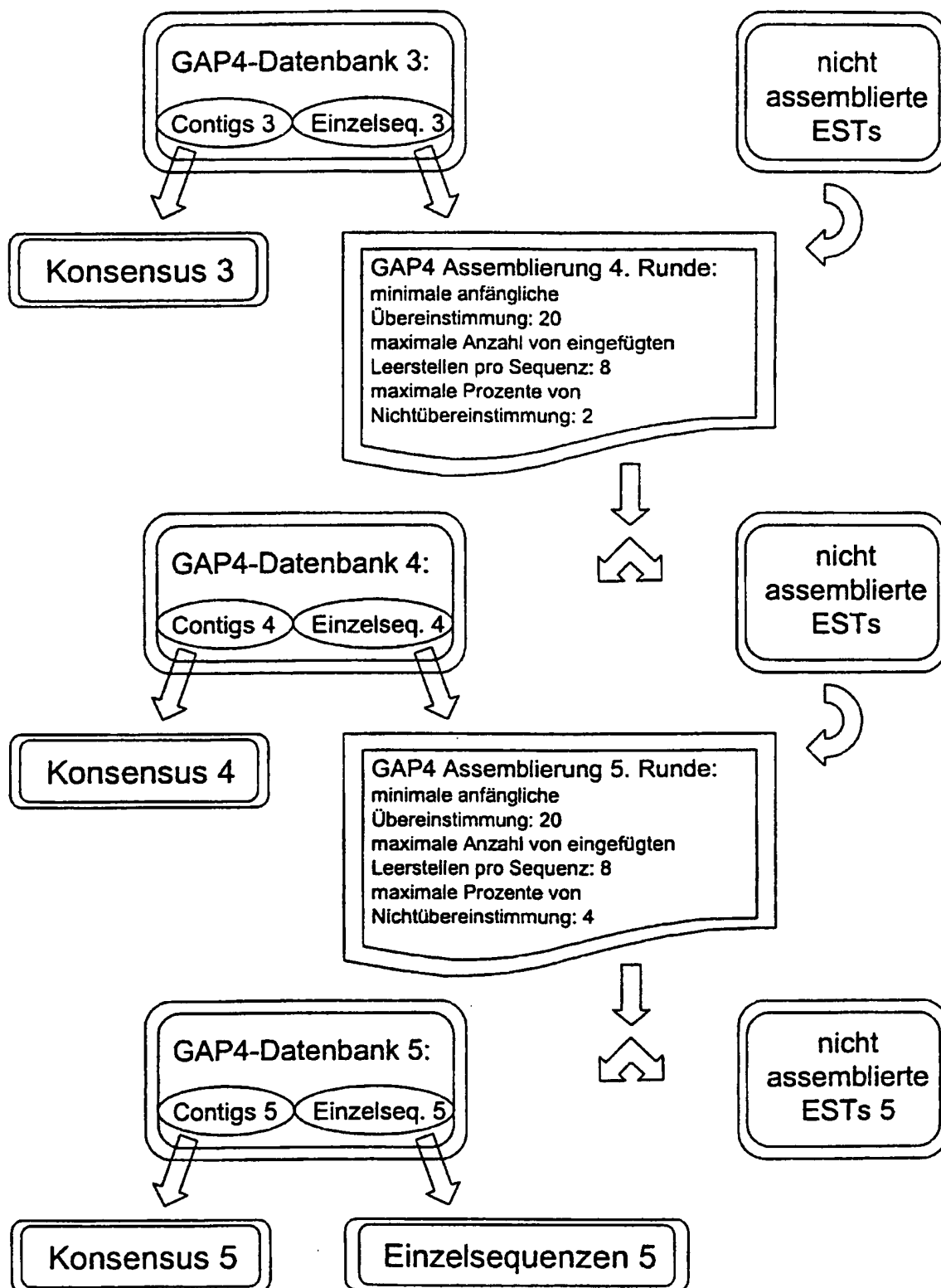


Fig. 2b2

5/10

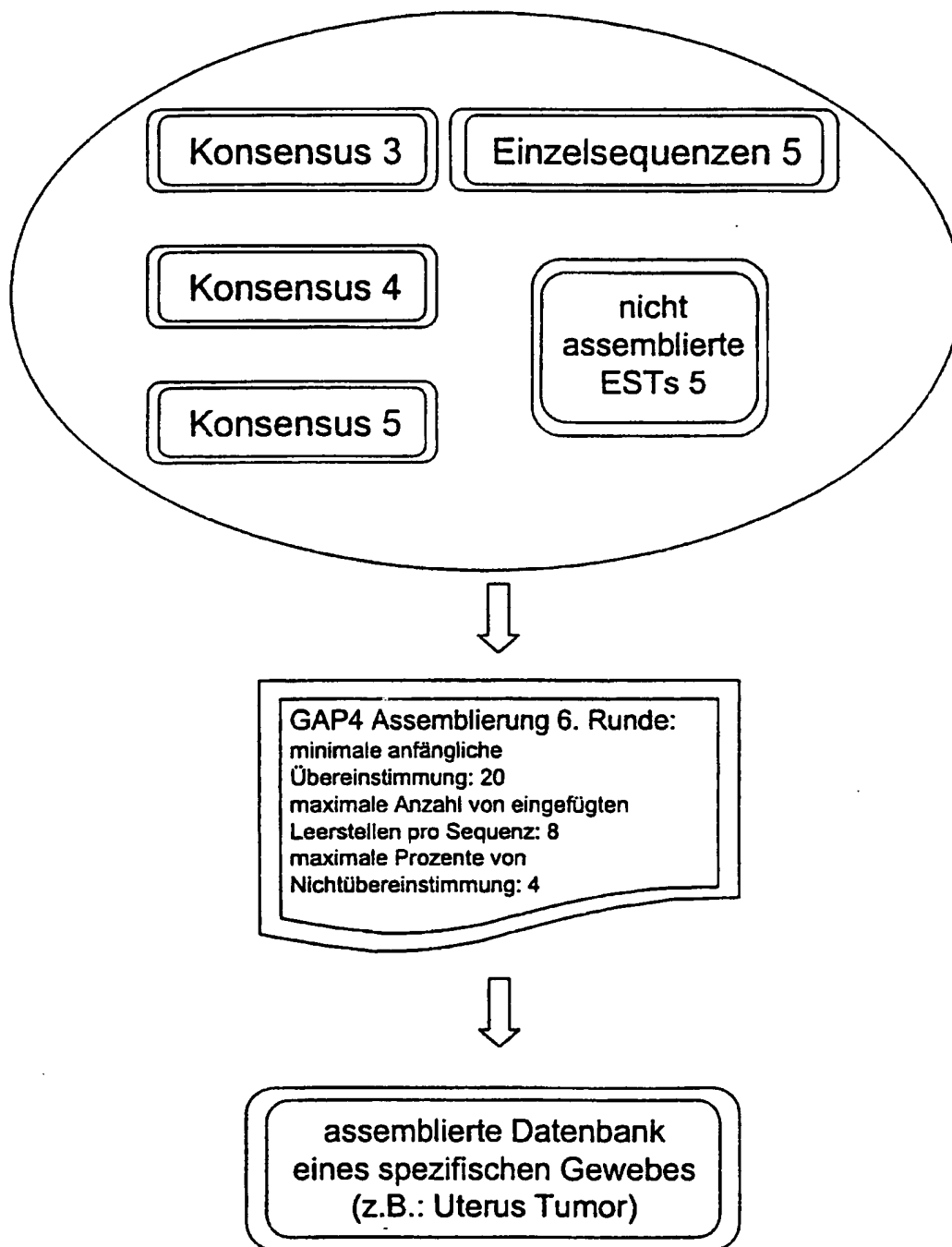


Fig. 2b3

6/10

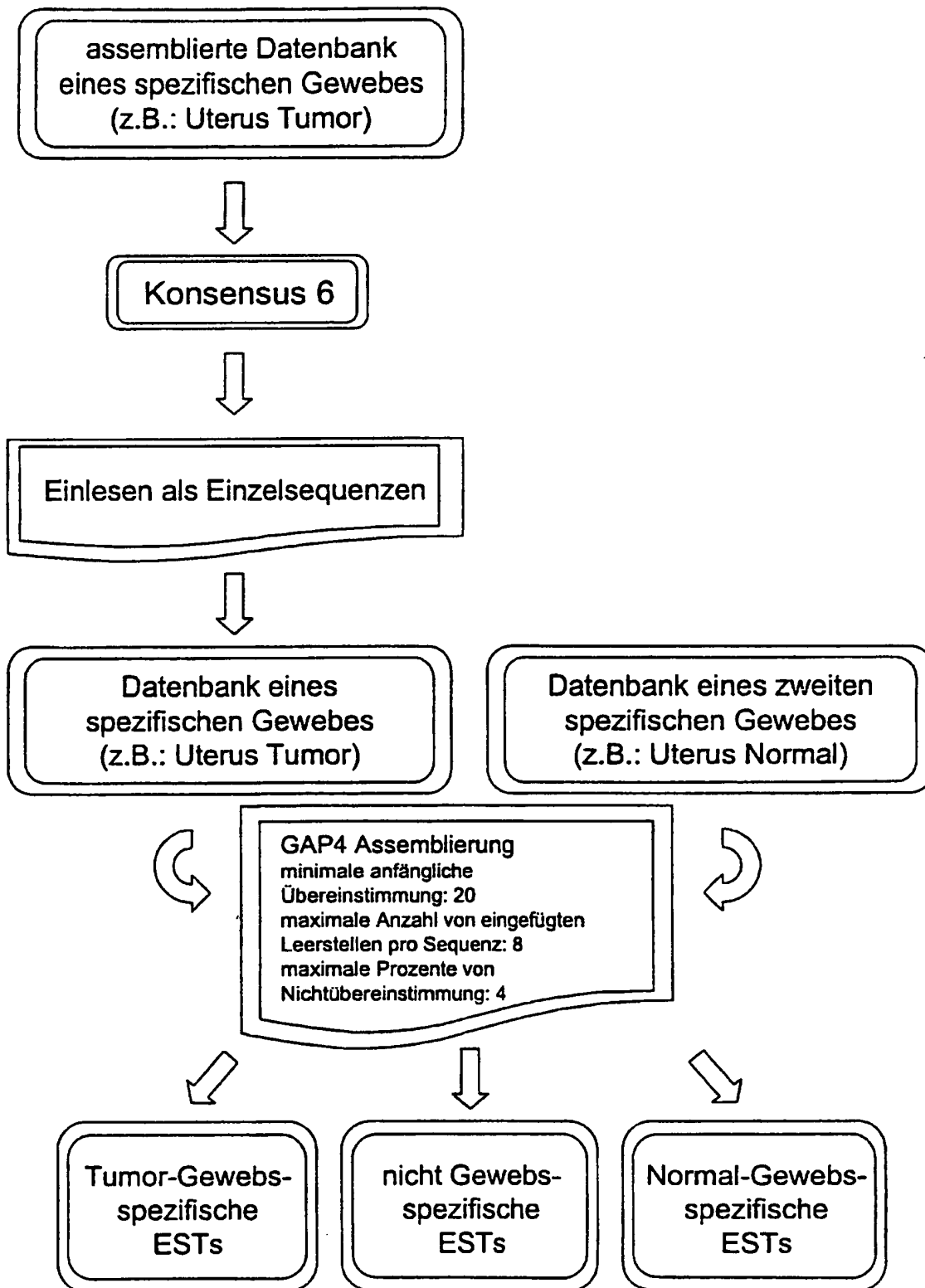


Fig. 2b4

7/10

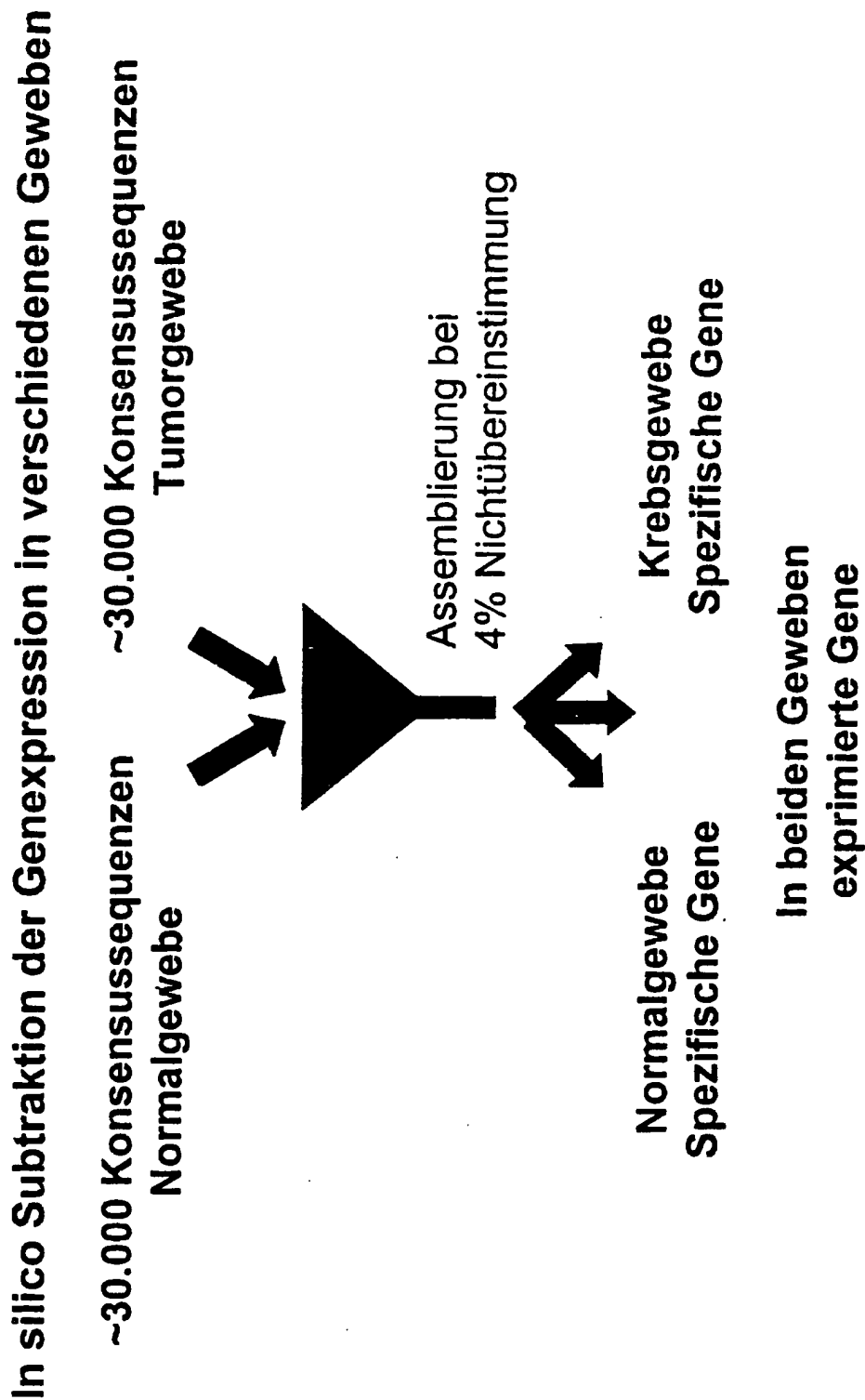


Fig. 3



8/10

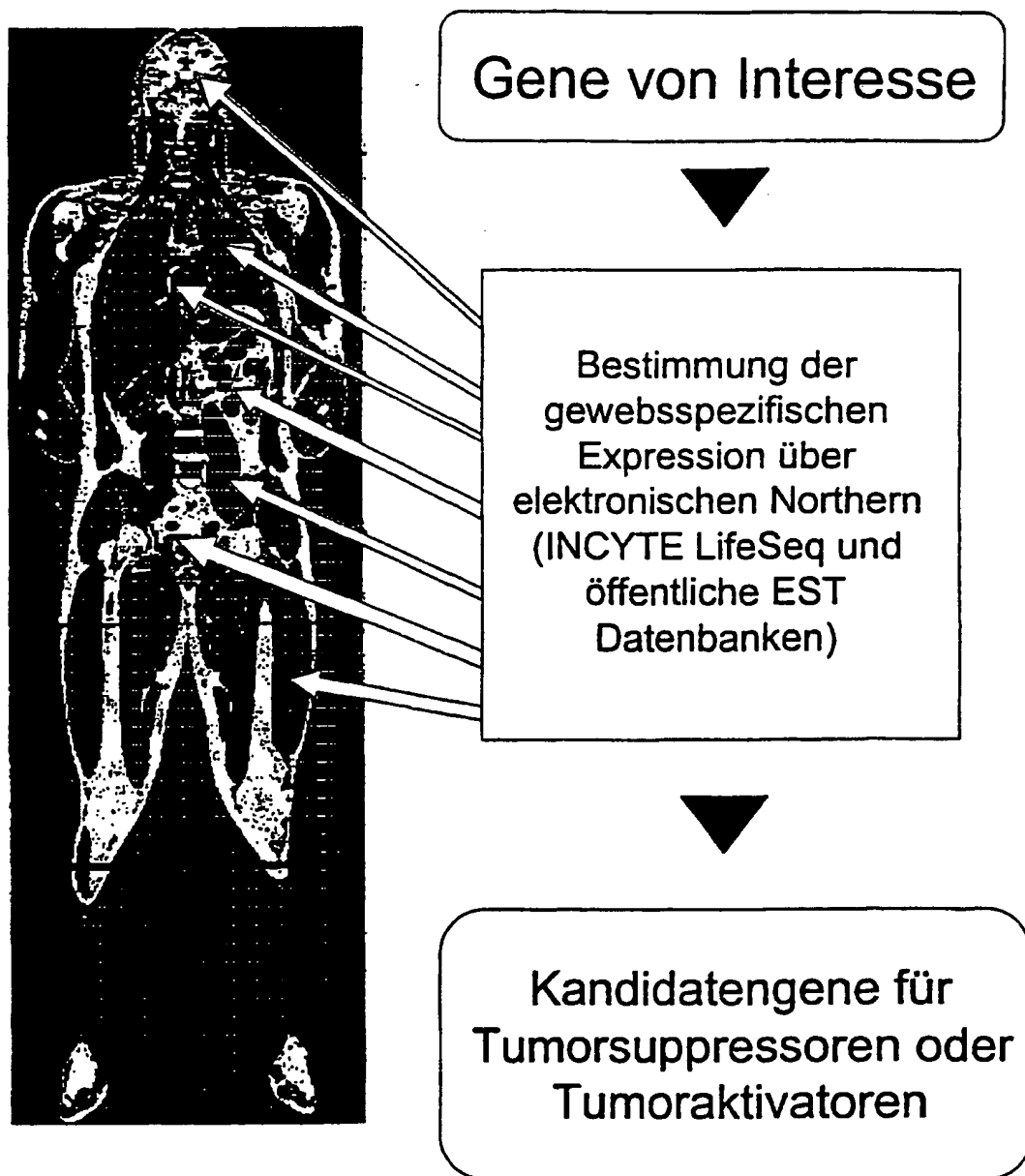


Fig. 4a

9/10

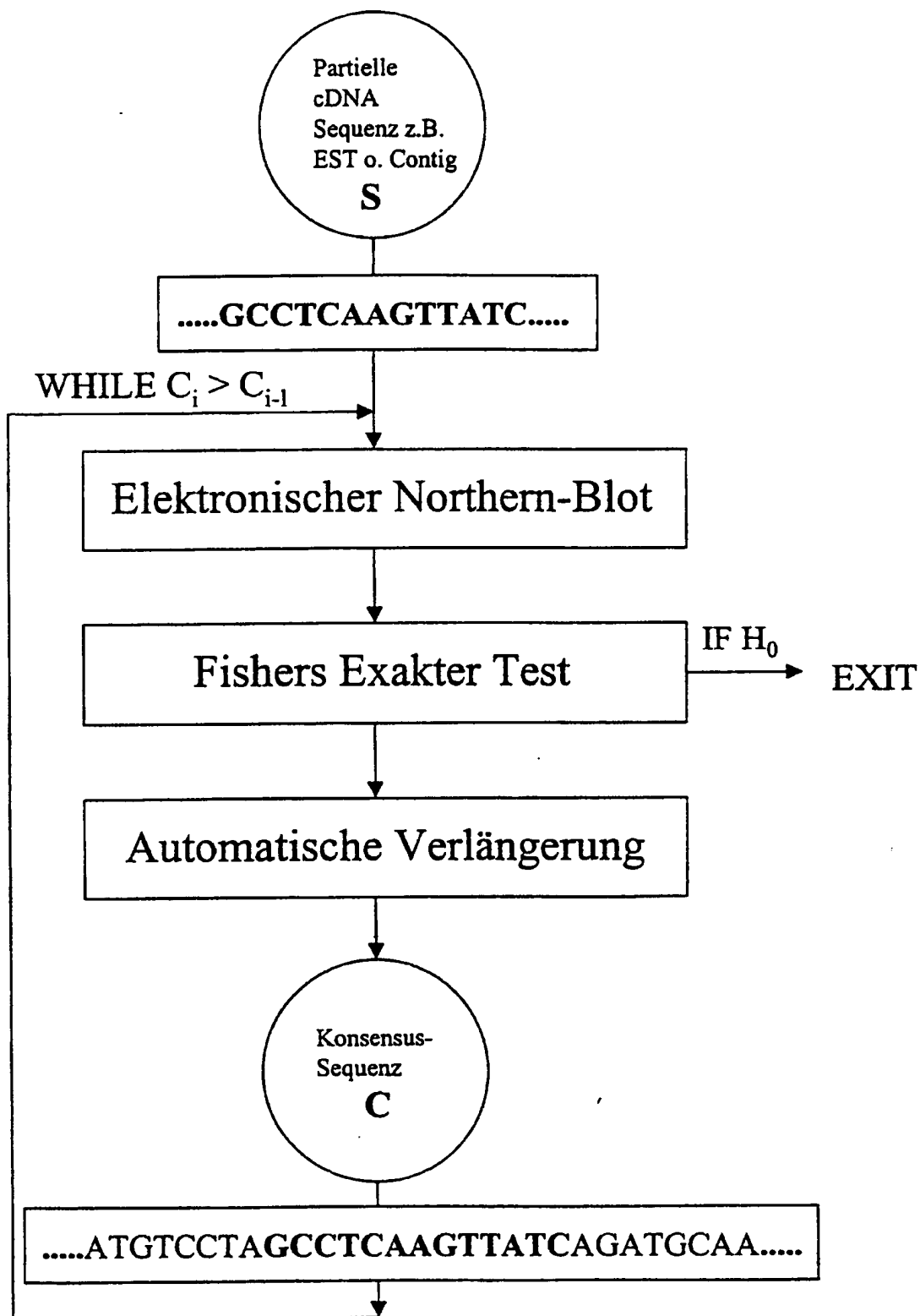


Fig. 4b

10/10

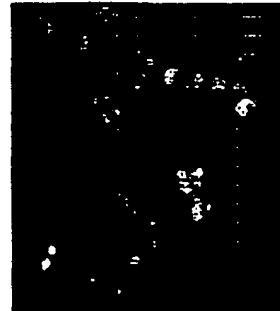
Isolieren von genomischen BAC und PAC Klonen



Chromosomale Klon-Lokalisation über FISH



Hybridisierungssignal



Sequenzierung von Klonen, die in Regionen lokalisiert sind, die chromosomale Deletionen in Prostata- und Brustkrebs aufweisen, führt zur Identifizierung von Kandidatengenen



Bestätigung der Kandidatengene durch Screening von Mutationen und/oder Deletionen in Krebsgeweben

Fig. 5